

UNIVERSIDAD DE GRANADA



TESIS DOCTORAL

**Análisis de Imágenes Funcionales Cerebrales
mediante Modelos de Mezcla de Gaussianas
y Mínimos Cuadrados Parciales para el
Diagnóstico de Alteraciones Neurológicas**

Noviembre, 2010

Fermín Segovia Román

Departamento de Arquitectura y Tecnología de Computadores

Editor: Editorial de la Universidad de Granada
Autor: Fermín Segovia Román
D.L.: GR 769-2012
ISBN: 978-84-694-1319-7

D. Juan Manuel Górriz Sáez, Doctor por la Universidad de Cádiz y la Universidad de Granada y Profesor Titular del Departamento de Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada y

D. Javier Ramírez Pérez de Inestrosa, Doctor por la Universidad de Granada y Profesor Titular del Departamento Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada,

MANIFIESTAN:

Que la presente Memoria titulada “*Análisis de Imágenes Funcionales Cerebrales mediante Modelos de Mezcla de Gaussianas y Mínimos Cuadrados Parciales para el Diagnóstico de Alteraciones Neurológicas*”, presentada por Fermín Segovia Román para optar al grado de Doctor Europeo por la Universidad de Granada, ha sido realizada bajo nuestra dirección. Con esta fecha, autorizamos la presentación de la misma.

Granada, a 10 de Noviembre de 2010

Fdo: Juan Manuel Górriz Sáez Fdo: Javier Ramírez Pérez de Inestrosa

Memoria presentada por Fermín Segovia Román para optar al Grado de Doctor por la Universidad de Granada.

Fdo: Fermín Segovia Román

Título de Doctor con Mención Europea

Con el fin de obtener la Mención Europea en el Título de Doctor (aprobada en Junta de Gobierno de la Universidad de Granada el 5 de Febrero de 2001), se han cumplido, en lo que atañe a esta Tesis Doctoral y a su Defensa, los siguientes requisitos:

- Durante su etapa de formación, el doctorando ha realizado una estancia de 3 meses fuera de España en una institución de enseñanza superior de otro Estado europeo cursando estudios o realizando trabajos de investigación que le han sido reconocidos por el órgano responsable del programa.
- Parte de la Tesis se ha redactado y presentado en una de las lenguas oficiales de la Unión Europea distinta a alguna de las lenguas oficiales en España. En concreto en Inglés.
- La Tesis ha sido informada por dos expertos pertenecientes a alguna institución de educación superior o instituto de investigación de un Estado miembro de la Unión Europea distinto de España.
- Un experto perteneciente a alguna institución de educación superior o instituto de investigación de un Estado miembro de la Unión Europea distinto de España, con el grado de doctor, y distinto de los mencionados antes, forma parte del tribunal evaluador de la Tesis.

*“Investigación es lo que hago
cuando no sé lo que estoy haciendo”*

Wernher von Braun

Agradecimientos

La realización de esta Tesis Doctoral ha sido un trabajo largo y complejo que no hubiera sido posible sin la ayuda de muchas personas de mi entorno que, con su motivación y apoyo, han sabido guiarme para conseguir este objetivo.

En primer lugar me gustaría agradecer a mis directores Juan Manuel Górriz y Javier Ramírez su dedicación y esfuerzo. Ellos me han demostrado que con trabajo y constancia se pueden conseguir grandes cosas. Gracias por enseñarme a INVESTIGAR. Quiero agradecer también su apoyo a Carlos G. Puntonet, sin el cual esta tesis habría estado mucho más cuesta arriba.

También me gustaría agradecer la buena disposición de mis compañeros del grupo SiPBA: Diego, Ignacio, Miriam, Rosa y Pablo. Gracias por tener siempre la mano tendida.

No menos importante ha sido el apoyo de mi familia. Gracias a mis padres y a mi hermana por compartir conmigo los momentos buenos y hacer vuestros también, los momentos malos.

No quisiera olvidarme de Jesús, Rosana y Fernando. Gracias por vuestro apoyo, sobre todo en la recta final de esta tesis, y por estar siempre dispuestos a echarme una mano.

Financiación

Las principales fuentes de financiación de esta Tesis Doctoral han sido las siguientes:

- **Beca de Formación de Investigadores del Plan Propio de Investigación de la Universidad de Granada**
Entidad financiadora: Universidad de Granada
Duración: 2008-2012
- **Detección de enfermedades neurológicas mediante clasificación y separación de señales** (Proyecto MEC PET2006-0253)
Entidad financiadora: Ministerio de Educación y Ciencia
Duración: 2007-2010
- **Diagnóstico avanzado de enfermedades neurológicas mediante técnicas de reconstrucción y modelado de imágenes tomográficas cerebrales** (Proyecto de excelencia TIC-02566)
Entidad financiadora: Junta de Andalucía
Duración: 2008-2012
- **Sistemas de ayuda al diagnóstico inteligente de enfermedades neurodegenerativas** (SADIEN)
Entidad financiadora: Universidad de Granada
Duración: 2008
- **Toolbox for Biomedical Signal and Image Processing based on Information Theory** (Ref.: HD2008-0029)
Entidad financiadora: Ministerio de Ciencia e Innovación
Duración: 2009-2010
- **Nuevas Técnicas de Reconstrucción, Procesado, Clasificación y Fusión de Imágenes Médicas para Diagnóstico Precoz de la Enfermedad de Alzheimer** (Proyecto TEC2008-02113/TEC)
Entidad financiadora: Ministerio de Ciencia e Innovación
Duración: 2009-2012
- **Plataforma abierta de procesamiento de imágenes para ayuda al diagnóstico de alteraciones neurológicas** (Proyecto de excelencia PAPI-ADAN, TIC-4530)
Entidad financiadora: Junta de Andalucía
Duración: 2010-2014

Resumen

En esta Tesis Doctoral se presentan dos nuevos métodos de análisis de imágenes funcionales que permiten mejorar la precisión de los sistemas de ayuda al diagnóstico de enfermedades neurodegenerativas tales como la enfermedad de Alzheimer (AD).

El primero de ellos hace uso del Modelo de Mezcla de Gaussianas (GMM) para agrupar o parcelar vóxeles similares. Para ello, en primer lugar, se construye una imagen denominada *imagen modelo*, que contiene las regiones de interés. Esta imagen puede ser definida de varias formas: como una media de imágenes normales o como la diferencia entre las imágenes normales y las imágenes patológicas. Posteriormente se aplica un algoritmo que modela esta imagen como una mezcla de Gaussianas y a partir de ella se genera un vector de características que “resume” cada imagen funcional. Este vector contiene un elemento por cada una de las Gaussianas de la mezcla y se calcula como la activación de la imagen completa para cada Gaussiana.

El segundo de los métodos desarrollado está basado en la técnica de Mínimos Cuadrados Parciales (PLS). Esta técnica descompone dos conjuntos de variables como el producto de dos matrices denominadas *scores* y *loadings* siguiendo un criterio de maximización de la covarianza. En nuestro caso, los dos conjuntos de variables son los formados por las imágenes funcionales y por las etiquetas de dichas imágenes. Tras la descomposición de estos conjuntos, la matriz de *scores* correspondiente al conjunto de las imágenes contiene proyecciones de los vóxeles que son obtenidas teniendo en cuenta las etiquetas de las imágenes. Estas proyecciones, conocidas como componentes PLS, son usadas como vectores de características.

Estos métodos han sido validados mediante la construcción de sistemas de ayuda al diagnóstico para AD consiguiendo resultados que superan ampliamente enfoques anteriores como los basados en la técnica de Voxel-As-Features y Análisis de Componentes Principales.

Por último, se ha implementado una herramienta software que permite llevar los últimos avances en sistemas CAD, como los presentados en esta Tesis, a los hospitales y centros médicos. Esta herramienta posee una interfaz gráfica intuitiva y fácil de usar, incluso por personas sin grandes conocimientos en informática.

ÍNDICE GENERAL

1. Introducción	1
1.1. Sistemas de ayuda al diagnóstico por ordenador	2
1.1.1. Modelos univariados	4
1.1.2. Modelos basados en la textura	5
1.1.3. Modelos basados en aprendizaje automático	7
1.2. Objetivos de esta tesis	9
1.3. Ejemplo de aplicación: La enfermedad de Alzheimer	10
1.3.1. Introducción	10
1.3.2. Diagnóstico de la enfermedad de Alzheimer	11
1.4. Estructura de la Tesis Doctoral	12
1.5. Publicaciones científicas	13
1.5.1. Artículos en revistas internacionales	13
1.5.2. Aportaciones a congresos internacionales	16
1.5.3. Capítulos de libros	23

I Técnicas de imagen médica usadas en el diagnóstico

de enfermedades neurodegenerativas	25
2. Técnicas tomográficas usadas en medicina nuclear	27
2.1. Introducción	28
2.2. Técnicas de neuroimagen estructural	28
2.3. Técnicas de neuroimagen funcional	29
2.3.1. SPECT	30
2.3.2. PET	31
2.3.3. Adquisición de las imágenes	33
2.3.4. Reconstrucción de las imágenes	34
2.4. Preprocesamiento de las imágenes	36
2.4.1. Normalización espacial	36
2.4.2. Normalización de intensidad	38
3. Bases de datos de imágenes	41
3.1. Base de datos del Hospital Virgen de las Nieves	42
3.2. Base de datos del centro PET Cartuja	43
3.3. Base de datos de la iniciativa ADNI	45
3.3.1. Protocolo de adquisición	45
3.3.2. Criterios de Etiquetado	48
4. La enfermedad de Alzheimer	51
4.1. La enfermedad de Alzheimer	52
4.1.1. Causas	53
4.1.2. Anatomía patológica	54
4.2. Técnicas de diagnóstico de AD	56
4.2.1. Examen del estado mental mínimo	57
4.2.2. Escala de deterioro global	57
4.2.3. Clasificación clínica de la demencia	59
4.2.4. Escala de evaluación para la enfermedad de Alzheimer	59
II Fundamentos teóricos	61
5. Métodos de clasificación	63
5.1. Introducción	64
5.2. Métodos Kernel	66
5.3. Clasificadores estadísticos	68
5.3.1. Clasificador de la media más cercana	69

5.3.2.	Clasificador lineal de Fisher	70
5.3.3.	Clasificador Bayesiano Ingenuo	70
5.3.4.	Análisis Discriminante Lineal	72
5.3.5.	Clasificador de los k Vecinos más Cercanos	72
5.3.6.	Arboles de Decisión	73
5.3.7.	Random Forests	74
5.3.8.	Máquinas de Vectores Soporte	79
5.4.	Extracción de características	88
5.5.	Evaluación de un clasificador	91
5.5.1.	Medidas de rendimiento	91
5.5.2.	Curva ROC	93
5.5.3.	Técnicas de validación cruzada	94
6.	Mapas Estadísticos Paramétricos	97
6.1.	Introducción	98
6.2.	Preprocesamiento en SPM	99
6.2.1.	Realineamiento	99
6.2.2.	Normalización espacial	100
6.2.3.	Normalización en intensidad	101
6.2.4.	Filtrado espacial	101
6.3.	Análisis estadísticos	102
6.4.	Mapa de estadísticos paramétricos	104
6.5.	Limitaciones de SPM	106
6.6.	Sistemas CAD basados en SPM	107
III	Aportaciones de la Tesis Doctoral	109
7.	Análisis de IF basado en el Modelo de Mezcla de Gaussianas	111
7.1.	Introducción	112
7.2.	Definición del modelo	113
7.3.	Estimación por máxima verosimilitud	115
7.4.	Algoritmo de maximización de la esperanza	117
7.4.1.	Algunos comentarios sobre el algoritmo EM	118
7.4.2.	Inicialización	118
7.4.3.	Selección del modelo	119
7.5.	Extracción de características	121

8. Análisis de IF basado en Mínimos Cuadrados Parciales	125
8.1. Introducción	126
8.2. Desarrollo matemático	127
8.3. Algoritmos PLS	130
8.3.1. NIPALS	130
8.3.2. SIMPLS	132
8.4. Extracción de características	133
9. Resultados experimentales	137
9.1. Descripción de los experimentos	138
9.2. Ajuste de parámetros de los clasificadores	139
9.3. Resultados para la base de datos HVN	140
9.3.1. Análisis de los resultados	151
9.4. Resultados para la base de datos CDPC	152
9.4.1. Análisis de los resultados	162
9.5. Resultados para la base de datos ADNI	163
9.5.1. Controles frente a Alzheimer	163
9.5.2. Controles y MCI frente a MCI converters y Alzheimer	174
9.6. Discusión	185
10.Herramienta para el estudio de imágenes médicas	189
10.1. Introducción	190
10.2. Formatos de archivo	191
10.3. Visualización y preprocesado de imágenes	192
10.4. Entrenamiento y clasificación	194
11.Conclusiones y trabajo futuro	197
11.1. Conclusiones	198
11.2. Trabajo futuro	200
IV Summary in English	201
12.Introduction	205
12.1. Computer-Aided Diagnosis systems	206
12.2. Goals of this Thesis	207

13. Analysis of FI based on the Gaussian Mixture Model	209
13.1. Introduction	210
13.2. GMMs for density estimation	211
13.3. Maximum Likelihood Estimation using the EM algorithm	211
13.4. The EM algorithm	213
13.4.1. Some remarks on the EM algorithm	214
13.4.2. Initialization	215
13.4.3. Model Selection	216
13.5. Feature Extraction	217
14. Analysis of FI based on the Partial Least Squares	221
14.1. Model description	222
14.2. SIMPLS algorithm	224
14.3. Feature extraction	224
15. Experiments and results	227
15.1. Image databases	228
15.1.1. HVN database	228
15.1.2. CDPC database	229
15.1.3. ADNI database	229
15.1.4. Image preprocessing	230
15.2. Experiments	231
15.3. Estimation of the classifier parameters	232
15.4. Results obtained using HVN database	233
15.5. Results obtained using CDPC database	233
15.6. Results obtained using ADNI database	235
15.7. Discussion	236
16. Tool for study of medical images	239
16.1. Introduction	240
16.2. File formats	240
16.3. Image Management	241
16.4. Training and classification	243
17. Conclusions and future work	245
17.1. Conclusions	246
17.2. Future work	248

Abreviaturas

AD:	Alzheimer's Disease (Enfermedad de Alzheimer)
ADAS:	Alzheimer's Disease Assessment Scale (Escala de Evaluación para la Enfermedad de Alzheimer)
ANN:	Artificial Neural Network (Redes Neuronales Artificiales)
BIC:	Bayesian Information Criterion (Criterio de Información Bayesiano)
CAD:	Computer-Aided Diagnosis (Diagnóstico Asistido por Ordenador)
CDR:	Clinical Dementia Rating (Clasificación Clínica de las Demencias)
CSF:	Cerebrospinal Fluid (Líquido Cefalorraquídeo)
CT:	Computer Tomography (Tomográfica Computarizada)
DT:	Decision Trees (Árboles de Decisión)
ECD:	EthylCysteinate Dimer (Dímero EtilicoCisteinato)
EM:	Expectation-Maximization (Maximización de la esperanza)
FBP:	Filtered Backprojection (Retroproyección Filtrada)
FDG:	Fludeoxyglucose (Fluorodesoxiglucosa)
FI:	Functional Imaging (Imagen Funcional)
FLD:	Fisher Linear Discriminant (Discriminador Lineal de Fisher)
fMRI:	Functional Magnetic Resonance Imaging (Imágenes Funcionales de Resonancia Magnética)
FTD:	Frontotemporal Dementia (Demencia Frontotemporal)
GDS:	Global Deterioration Scale (Escala de Deterioro Global)
GM:	Grey Matter (Materia Gris)
GMM:	Gaussian Mixture Model (Modelo de Mezcla de Gaussinas)
HMPAO:	HexaMethylPropyleneAmine Oxime (HexaMetiloPropilenoAmina Oxima)
ICA:	Independent Component Analysis (Análisis de Componentes Independientes)
IF:	Imagen Funcional
KNN:	k Nearest Neighbor (k Vecinos más Cercanos)
LDA:	Linear Discriminant Analysis (Análisis Discriminante Lineal)

MCI:	Mild Cognitive Impairment (Deterioro Cognitivo Leve)
MLE:	Maximum Likelihood Estimation (Estimación por Máxima Verosimilitud)
MMSE:	Mini-Mental State Examination (Examen del Estado Mental Mínimo)
MRI:	Magnetic Resonance Imaging (Imágenes de Resonancia Magnética)
NB:	Naive Bayes (Clasificador Bayesiano Ingenuo)
NM:	Nearest Mean (Media más cercana)
PCA:	Principal Component Analysis (Análisis de Componentes Principales)
PCR:	Principal Component Regression (Regresión de Componentes Principales)
PET:	Positron Emission Tomography (Tomografía por Emisión de Positrones)
PLS:	Partial Least Squares (Mínimos Cuadrados Parciales)
RF:	Random Forests
ROI:	Region Of Interest (Región de Interés)
RR:	Ridge Regression (Regresión Contraída)
SLT:	Statistical Learning Theory (Teoría de Aprendizaje Estadístico)
SPECT:	Single Photon Emission Computed Tomography (Tomografía Computarizada por Emisión de un Solo Fotón)
SPM:	Statistical Parametric Mapping (Mapas Estadísticos Paramétricos)
SVM:	Support Vector Machine (Maquinas de Vectores Soporte)
VAF:	Voxels-As-Features (Vóxeles Como Características)
WM:	White Matter (Materia Blanca)

CAPÍTULO

1

INTRODUCCIÓN

La demencia es uno de los trastornos neurodegenerativos más frecuentes entre las personas de edad avanzada y conlleva consecuencias dramáticas para la salud e importantes implicaciones socio-económicas [Ritchie & Lovestone, 2002]. Con el envejecimiento de la población en los países desarrollados, la relevancia de esta enfermedad aumenta cada día. En los últimos años, el uso de sistemas informáticos basados en imágenes tomográficas del cerebro, ha aportado mejoras significativas en el diagnóstico y seguimiento de diversas modalidades de demencia. Sin embargo el tratamiento informático de las imágenes tomográficas no está exento de problemas y tiene aún un gran margen de mejora.

1.1. Sistemas de ayuda al diagnóstico por ordenador

Varias investigaciones recientes [Thompson & Apostolova, 2007; Whitwell *et al.*, 2007; Johnson *et al.*, 2006] sugieren que el proceso neurodegenerativo asociado a la demencia comienza varios años de que los pacientes tengan síntomas, que es cuando se pueden usar los test cognitivos para diagnosticar la enfermedad. Con el objetivo de anticipar el diagnóstico y mejorar de esta forma los tratamientos, las técnicas de imagen médica basadas en medicina nuclear suponen un avance significativo. Estas técnicas permiten obtener mapas tridimensionales de la actividad cerebral para, a partir de ellos, emitir un diagnóstico en función de cambios neurodegenerativos que se hayan podido producir en el cerebro.

La tomografía computarizada por emisión de un solo fotón (SPECT) y la tomografía por emisión de positrones (PET) son dos modalidades de imágenes no invasivas ampliamente usadas en medicina para el diagnóstico y seguimiento de varias dolencias [Miller *et al.*, 1991; Starkstein *et al.*, 1994; Delbeke *et al.*, 1999; Eary, 1999]. Los mapas de la actividad cerebral se obtienen de forma similar con estas dos técnicas. En primer lugar se inyecta al paciente un medicamento conocido como radiofármaco que contiene un isótopo radioactivo. Pasado un tiempo (de minutos u horas) el radiofármaco comienza a descomponerse emitiendo partículas que son detectables desde el exterior y que permiten saber cómo se ha distribuido el radiofármaco en el interior del cuerpo. Más concretamente, para un SPECT cerebral se suele usar como radiofármaco el $^{99}\text{Tc-ECD}$. Esta sustancia se mantiene en el torrente sanguíneo durante su descomposición radioactiva en la que emite rayos gamma. Con una cámara gamma colocada alrededor de la cabeza se puede medir el riego sanguíneo del cerebro que será proporcional al consumo de oxígeno de las neuronas. Con esta información y con ayuda de un ordenador, es posible construir un mapa de la actividad cerebral del individuo. Para PET, en cambio, se suele usar el radiofármaco $^{18}\text{F-FDG}$ que las células confunden con glucosa y que, por tanto, permite medir el consumo de glucosa de las células. Al igual que el consumo de oxígeno, el consumo de glucosa es indicativo de la actividad cerebral.

La capacidad de predicción de la medicina nuclear con respecto a las enfermedades neurodegenerativas ha sido avalada por un elevado número de estudios [Hoffman *et al.*, 2000; Silverman *et al.*, 2001a; Higdon *et al.*, 2004]. En 1994 el Subcomité de Normas de Calidad de la Academia Americana de Neurología publicó los “parámetros prácticos” para el diagnóstico y evaluación de personas que pudieran padecer alguna demencia [Chui & Zhang, 1997]. Estas directri-

ces establecían el examen neurológico como el test estándar y el procedimiento esencial del proceso de diagnóstico. Sin embargo, los estudios de neuroimagen eran sugeridos sólo como opción y definidos como “estrategias de certeza clínica incierta”. Pocos años después, en 1997, la utilidad de estos parámetros prácticos para la evaluación de la demencia se examinó mediante la revisión de un gran número de casos consecutivos. El principal resultado de estas evaluaciones es que los estudios de neuroimagen mejoraban la precisión en el diagnóstico, aunque con un aumento previsible de los costes. Los resultados demostraron que los estudios de neuroimagen cambian con frecuencia el diagnóstico clínico (entre el 19 % y el 28 % de los casos) y el tratamiento de los pacientes (15 %).

En la última década, se ha generalizado el uso de imágenes SPECT y PET para el diagnóstico de diferentes tipos de demencia. Habitualmente, estas imágenes son valoradas por expertos que evalúan visualmente la presencia de las características típicas de la enfermedad. Sin embargo, la exploración visual de las imágenes sigue siendo un proceso subjetivo que requiere médicos experimentados. Por este motivo, el uso de sistemas informáticos que ayuden a los médicos en la manipulación y análisis de las imágenes tomográficas supone una mejora notable en la precisión del diagnóstico precoz de muchas enfermedades neurodegenerativas. El uso de estos sistemas, conocidos como sistemas CAD (del inglés Computer-Aided Diagnosis Systems), persigue tres objetivos fundamentalmente:

- Aprovechar toda la información que aportan las imágenes (diferencias muy pequeñas en la intensidad de los vóxeles son inapreciables para el ojo humano).
- Eliminar la subjetividad inherente a la exploración visual de las imágenes.
- Aportar un juicio de valor sobre una imagen (basado en la comparación con otras imágenes) que ayude al equipo médico en la toma de decisiones.

Existen diferentes enfoques para el tratamiento de las imágenes en un sistema CAD. El enfoque más familiar para los investigadores de neuroimagen es el enfoque univariado según el cual las imágenes son procesadas vóxel a vóxel. El enfoque multivariado, en cambio, trata un grupo de vóxeles vecinos como un solo dato. En este caso existen dos variantes, los modelos basados en la textura y los modelos basados en aprendizaje automático, en los que usa una imagen cerebral completa como una observación. En los siguientes apartados se describen estos enfoques.

1.1.1. Modelos univariados

En principio, el metabolismo anormal del cerebro, característico en las demencias, podría ser detectado con un modelo estadístico que compare neuroimágenes de un solo paciente tomadas en condiciones *neurologicamente distintas*. Esto requiere realizar diferentes mediciones en el tiempo, para medir la funcionalidad cerebral del paciente en diferentes etapas (desde que estaba sano hasta fases avanzadas de la enfermedad), lo cual supone demasiado tiempo. No obstante, se puede llevar a cabo una comparación similar siempre y cuando contemos con un número suficientemente grande de pacientes *neurologicamente normales* y su metabolismo cerebral sea medido con el mismo procedimiento/sistema que el paciente que sufre la enfermedad. En este caso podemos suponer que el paciente patológico y los sujetos *normales* son la misma persona bajo diferentes condiciones: la condición patológica (en una medida) y la condición sana (en varias medidas). Este enfoque requiere:

- Datos de referencia apropiados, esto es, debidamente validados como *normales* y en número suficiente para la comparación.
- Algoritmos de registro espacial que eliminen las diferencias de forma y tamaño debidas al hecho de que las imágenes proceden de distintas personas.
- Test estadísticos que tengan en cuenta las variables neurológicas de interés que afectan a la hipótesis estadística (por ejemplo, las diferentes edades o sexos de los sujetos).
- Presentación adecuada de los resultados, mostrando regiones de hipoperfusión o hipometabolismo sobre un atlas conocido del cerebro y mediante el uso de técnicas de renderizado 3D.

A principios del 2000 se demostró [Signorini *et al.*, 1999] que un modelo basado en Mapas Estadísticos Paramétricos (SPM) ¹, podría ser un modelo adecuado para describir el patrón neurodegenerativo del funcionamiento cerebral. Sin embargo, los modelos basados en SPM, al igual que todos los enfoques univariados, tienen el inconveniente de que cada vóxel es tratado independientemente, y por lo tanto no se tienen en cuenta las relaciones entre distintas regiones del cerebro. A pesar de esto, SPM se utiliza frecuentemente para ayudar en el diagnóstico neurológico y su uso se ha convertido en una práctica estándar para la evaluación neurológica.

¹SPM será descrito ampliamente en el capítulo 6

Recientemente, la suma de los valores t obtenidos por SPM en regiones que son típicamente hipometabólicas en enfermos de Alzheimer se ha utilizado como un indicador de la enfermedad, con una precisión del 93% [Herholz *et al.*, 2002]. Además de SPM, en los últimos años se han desarrollado otros modelos estadísticos basados en ordenador que proporcionan nuevos enfoques a nivel de vóxel. Por ejemplo, el paquete de software comercial NEUROSTAT & 3D-SSP se ha utilizado con éxito para identificar alteraciones metabólicas en la demencia por deterioro cognitivo leve (MCI) [Hosaka *et al.*, 2002]. Los modelos univariados en general se basan en la detección de los vóxeles anormales o en el reconocimiento automático de la distribución anatómica de las anomalías metabólicas en sujetos patológicos pero no realizan una comparación con sujetos *normales*. Esto tiene el inconveniente de que los usuarios del sistema tienen que comprobar la validez de sus resultados, comparándolos con los datos de sujetos *normales*. Otro estudio realizado recientemente demostró que un análisis vóxel a vóxel, de ^{18}F FDG-PET aumenta la precisión en el diagnóstico tanto de la enfermedad de Alzheimer (AD, del inglés Alzheimer's Disease) como de la demencia frontotemporal (FTD), en particular cuando los resultados de una evaluación clínica no son claros [Foster *et al.*, 2007].

1.1.2. Modelos basados en la textura

La textura es la variación en los datos a escalas más pequeñas que la escala de interés [Petrou, 2010]. De acuerdo con esta definición, incluso las variaciones debidas al ruido pueden considerarse como una textura. Sin embargo, para los investigadores de procesamiento de imágenes y reconocimiento de patrones, la textura es una propiedad intrínseca de una superficie reflejada en una imagen y no algo causado por el instrumento con el que se ha capturado la imagen, como sí lo es el ruido. Una vez que aceptamos que una textura es una propiedad intrínseca del objeto fotografiado, la textura se convierte en un dato valioso para el reconocimiento de este objeto. La característica más importante de la textura es que depende de la escala: diferentes tipos de textura son visibles en diferentes escalas. Con el fin de poder utilizar la textura para identificar tipos de tejido u órganos humanos, debemos ser capaces de medirla de forma reproducible.

Las texturas se han utilizado frecuentemente en aplicaciones biomédicas. Por ejemplo, en mamografía se ha utilizado tanto para contribuir al reconocimiento del tipo de tejido como para ayudar al registro de la imagen [Petroudi & Brady, 2004; C.J.Rose, 2005], a la clasificación de patrones mamográficos [Petroudi *et al.*, 2003; Petroudi & Brady, 2004; Tao *et al.*, 2008] y a la caracterización de la densidad de la mama [Bovis & Singh, 2002; Oliver *et al.*, 2005; Muhimmah

& Zwiggelaar, 2006]. En general, se pueden distinguir dos grandes enfoques, a saber, la segmentación del tejido, y la caracterización de los tejidos ya segmentados, que se supone que tienen una textura uniforme. El segundo de ellos se trata en la mayoría de los artículos sobre este tema y es especialmente relevante para caracterizar enfermedades que afectan a todo el cerebro. El primer enfoque se ha utilizado sobre todo para segmentar la degeneración del tejido local, como por ejemplo, el crecimiento de un tumor. Un primer trabajo en la caracterización de los estados cerebrales degenerativos fue presentado en [Freeborough & Fox, 1998]. En [Sayeed *et al.*, 2002], los autores utilizan características basadas en la textura y calculadas directamente de las imágenes PET para distinguir entre pacientes de AD y controles normales. La idea había venido de un estudio anterior, que había demostrado que los cerebros de los pacientes con trastornos degenerativos tienden a ser más isotrópicos (tienen propiedades físicas uniformes en todas direcciones) que los cerebros sanos [Kovalev & Petrou, 1996; Kovalev *et al.*, 1999]. En [Kovalev *et al.*, 2003] las medidas de la isotropía de la textura se utilizaron para caracterizar los cerebros de pacientes esquizofrénicos a partir de datos de resonancia magnética. Sus autores demostraron que en las escalas de resolución de 1 ó 2 mm, el cuarto inferior de los cerebros de pacientes esquizofrénicos muestra diferencias estructurales significativas en la materia gris respecto a los cerebros de controles normales. En [Segovia-Martinez *et al.*, 2001] se demostró que existe una fuerte correlación entre las características isotrópicas de la textura, calculadas a partir de los datos de resonancia magnética, y la puntuación de los pacientes (y los controles normales) en el Examen del Estado Mental Mínimo (MMSE). Este es un resultado notable, ya que muestra una correlación entre una medida pura de procesamiento de la imagen y la puntuación en un test psicológico.

Dejando a un lado la caracterización de los tejidos segmentados manualmente y situándonos en el problema de la segmentación del tejido, la idea de utilizar las estadísticas de orden superior para caracterizar y segmentar tejidos se hizo posible con la disponibilidad de datos en 3D de alta resolución. En [Petrou *et al.*, 2006] los autores fueron capaces de mostrar que los límites presentes en los datos, pero invisibles al ojo humano, podían convertirse en visibles mediante el cálculo de la asimetría dentro de cada mitad de una ventana deslizante. Los autores, utilizando este enfoque, identificaron los límites alrededor de los glioblastomas (tumores malignos del sistema nervioso central). Más recientemente, en [ans S. Bose *et al.*, submitted 2010], los autores separaron pacientes con síntomas de la enfermedad de Huntington y controles normales, utilizando estadísticas de orden superior calculadas a partir de capas segmentadas del hipocampo. Para mejorar la exactitud de estas estadísticas se utilizó un método de estimación

basado en *kernels* (véase sección 5.2).

Aunque la textura se ha utilizado en cientos de artículos sobre el procesamiento de imágenes médicas, el potencial de estos métodos no ha sido plenamente explotado todavía. Una dirección de investigación muy prometedora son las estadísticas de orden superior, ya que se sabe que el ojo humano tiene dificultades para discernir las transiciones en estas estadísticas, y el uso de dicho análisis puede ayudar a revelar características de los datos que de otro modo no son visibles. Sin embargo, esto plantea un problema: si no podemos ver estos cambios en los datos, ¿cómo estamos seguros de que lo que calculamos está realmente ahí? La respuesta está en los tests de consistencia que se hicieron en [Petrou *et al.*, 2006], si bien, una respuesta realmente fiable sólo será posible si se llevan a cabo estudios que sigan a los pacientes durante unos años, y los resultados del análisis post-muerte se comparan con los datos del análisis de imágenes de la historia clínica del paciente. Tales estudios son muy difíciles, pero son el camino a seguir. Hay que tener cuidado, sin embargo, sobre el uso de métodos basados en la textura ya que la variación en los datos observados puede deberse a las imperfecciones en el proceso de captura más que a la variación del tejido original. De hecho, la ausencia de homogeneidad del campo magnético durante la captura de una MRI es sobradamente conocida. El único supuesto en que las conclusiones basadas en el análisis de textura son totalmente fiables es cuando los datos son capturados por el mismo instrumento y bajo el mismo protocolo, ya que la ausencia de homogeneidad del aparato de captura será la misma para todas los datos incluidos en el análisis.

1.1.3. Modelos basados en aprendizaje automático

La mayor parte de los sistemas CAD desarrollados en los últimos años están basados en sistemas multivariados que toman una imagen cerebral completa como una única observación. A pesar de que los datos para la clasificación se determinan a partir de la imagen cerebral completa, muchos de estos sistemas usan sólo un conjunto reducido de vóxeles o regiones para distinguir entre imágenes patológicas y controles. El auge de los sistemas multivariados se debe, en gran parte, a los avances recientes en la definición de los clasificadores estadísticos que hacen posible la construcción de clasificadores más fiables y con mayor capacidad de generalización, capaces de afrontar mejor el problema del pequeño tamaño muestral [Duin, 2000]. En este sentido, las Máquinas de Vectores Soporte (SVM) han centrado recientemente la atención de la comunidad de reconocimiento de patrones, debido a una serie de ventajas teóricas y computacionales derivadas de la Teoría del Aprendizaje Estadístico (SLT) [Vapnik,

1995, 1998].

Varios modelos basados en SVM se han utilizado recientemente para ayudar a distinguir entre enfermos de AD y controles de avanzada edad a partir de imágenes MRI [Fan *et al.*, 2008; Vemuri *et al.*, 2008]. Igualmente, en [Magnin *et al.*, 2009] se muestra un modelo basado en SVM para distinguir entre enfermos de AD y sujetos sanos de edad avanzada usando el cerebro completo a partir de imágenes MRI. En este trabajo se realizan varios análisis de histogramas a partir de las imágenes, con el fin de extraer las características teniendo en cuenta la distribución de la materia gris (GM), la materia blanca (WM), y el líquido cefalorraquídeo (CSF). El método propuesto *parcela* el cerebro de los pacientes creando regiones de interés (ROIs) y para cada ROI se realiza un análisis del histograma con la distribución de la intensidad de los vóxeles. Este análisis permite calcular un parámetro que identifica las respectivas contribuciones de GM, WM y CSF en cada región. Finalmente, usando este parámetro, que es estimado en todas las ROI, se lleva a cabo la clasificación mediante SVM. Otra aplicación de un modelo basado en SVM aplicado sobre imágenes MRI se presenta en [Klöppel *et al.*, 2008]. Los objetivos de este estudio fueron evaluar el grado de éxito de SVM asignando diagnósticos individuales, y determinar si la combinación de datos procedentes de diferentes cámaras y diferentes centros obtiene resultados de clasificación válidos. Usando SVM lineal, el método propuesto clasificó correctamente el 89 % de los pacientes (confirmado mediante autopsia) en dos clases: AD y degeneración lobular frontotemporal. Según sus autores, este trabajo constata que el diagnóstico por ordenador es igual o mejor que el alcanzado por los radiólogos [Klöppel *et al.*, 2008]. Por otro lado, un método basado en SVM no lineal junto con las medidas morfológicas de los vóxeles fue propuesto en [Lao *et al.*, 2004]. Estas medidas se obtienen a partir de las transformaciones que igualan la imagen a una plantilla (normalización espacial). De esta forma se pueden detectar patrones sutiles y espacialmente complejos que a menudo no son detectados por los métodos basados en vóxeles. También se han utilizado modelos basados en SVM para el análisis y clasificación de otras modalidades de imagen médica como fMRI [ao Miranda *et al.*, 2005; ao Ricardo Sato *et al.*, 2009], SPECT [López *et al.*, 2009a; Fung & Stoeckel, 2007] y PET [Hinrichs *et al.*, 2009]. Uno de los métodos más simples para clasificar imágenes funcionales usando técnicas de aprendizaje supervisado es el conocido como VAF (Voxels-As-Features) [Stoeckel *et al.*, 2001]. Este método separa pacientes de AD y controles a partir de imágenes SPECT. Para ello usa un clasificador SVM entrenado con todos los vóxeles que superan un cierto umbral de intensidad. Lo más destacado de este método es que, a pesar de su simplicidad, consigue resultados similares a otros métodos mucho más sofisticados.

Además de las Máquinas de Vectores Soporte, las Redes Neuronales Artificiales (ANN, del inglés Artificial Neuronal Network) suponen otro enfoque multivariado que proporciona elevadas tasas de sensibilidad y especificidad en la separación de imágenes. Una ANN puede definirse como un sistema de procesamiento de información distribuido en paralelo que puede adquirir, almacenar y utilizar conocimiento experimental. Al igual que SVM es un método de aprendizaje supervisado que utiliza un procedimiento de entrenamiento para aprender las diferencias entre las clases que se quieren clasificar. En [Acton & Newberg, 2006] se presenta un método basado en ANN para la clasificación de controles y enfermos de Parkinson que consigue elevadas tasas de sensibilidad y especificidad. Otro trabajo desarrollado recientemente usa ANN para discriminar pacientes esquizofrénicos y controles mediante imágenes ^{18}F -DOPA-PET [Bose *et al.*, 2008]. Este modelo clasifica correctamente el 94 % de los controles y el 89 % de los pacientes, lo que equivale a obtener una sensibilidad del 89 % y una especificidad del 94 %.

1.2. Objetivos de esta tesis

Esta tesis doctoral tiene como propósito mejorar los métodos de modelado y análisis de las imágenes tomográficas usadas en neurología para el diagnóstico de enfermedades neurodegenerativas como la demencia. Más concretamente los objetivos de este trabajo pueden dividirse en dos vertientes:

- Por un lado, mejorar los algoritmos de análisis de imágenes tomográficas en los que están basados los sistemas de ayuda al diagnóstico. El trabajo se centrará en los sistemas multivariados, que toman una imagen cerebral completa como una observación, debido a su mayor potencial, y a que no requieren que el usuario interprete el resultado de los análisis. Estos sistemas hacen uso de técnicas de aprendizaje estadístico supervisado para emitir un juicio de valor sobre una imagen funcional del cerebro de un paciente. La mejora de estos algoritmos pasa ineludiblemente por el desarrollo de métodos que eviten el problema del pequeño tamaño muestral, que sufren enfoques anteriores. Los nuevos métodos propuestos se derivarán del estudio de dos modelos distintos. El primero de ellos, el Modelo de Mezcla de Gaussianas, es un modelo clásico de *clustering* que puede usarse para agrupar o *parcelar* vóxeles similares de una imagen tomográfica. El segundo, el modelo de Mínimos Cuadrados Parciales, es un modelo similar al Análisis de Componentes Principales que ha sido usado
-

satisfactoriamente para regresión y que puede usarse también para tareas de reducción de la dimensionalidad. El desarrollo de sistemas CAD más precisos y eficientes tiene como fin último anticipar la detección de las enfermedades neurológicas permitiendo anticipar también el tratamiento y con ello mejorar y alargar la vida del paciente. Dado que el aprendizaje del sistema estará basado en un etiquetado previo realizado por expertos médicos, éste se limitará a reproducir con la máxima precisión posible el conocimiento de estos expertos.

- Por otro lado, la implementación de una herramienta informática que facilite el acceso por parte del personal sanitario a los métodos desarrollados en éste y otros trabajos de investigación similares. Esta herramienta debe cubrir todo el procesamiento que ha de seguir la imagen funcional desde que sale de la cámara hasta que es válida para el diagnóstico. Asimismo su interfaz debe ser clara y sencilla de forma que pueda ser utilizada por personas sin grandes conocimientos en el procesamiento de imágenes.

1.3. Ejemplo de aplicación: La enfermedad de Alzheimer

Una de las enfermedades neurodegenerativas que se verán beneficiadas por las mejoras derivadas del uso de los métodos de análisis de imágenes tomográficas presentados en esta tesis, es la enfermedad de Alzheimer. Por este motivo, dichos métodos han sido validados mediante el desarrollo de diferentes sistemas CAD para AD.

1.3.1. Introducción

La enfermedad de Alzheimer es una enfermedad neurodegenerativa que en una primera fase afecta a la memoria pero que gradualmente va deteriorando todas las funciones cognitivas y conductuales causando incluso la muerte. La esperanza de vida de un enfermo de Alzheimer es de, aproximadamente, 10 años tras el diagnóstico [Munoz, 2003] aunque esto puede variar en función del tratamiento que se siga y del estado de la enfermedad en el momento del diagnóstico. Un diagnóstico precoz es fundamental para frenar el avance de la enfermedad alargando la vida del paciente y mejorando su calidad de vida.

Considerada durante muchos años como una enfermedad rara, afecta a entre el 5 y el 7% de las personas de más de 65 años y al 80% de los mayores de 85

años. En números absolutos, la enfermedad de Alzheimer afecta a 30 millones de personas en todo el mundo, de los cuales 8 millones son europeos y 700.000 residen en España. Millones de familiares de enfermos ven su vida trastornada por culpa de la enfermedad. Debido al envejecimiento de la población en los países desarrollados, se prevé que el número de enfermos de AD se duplique en 2020 y se triplique en 2050 [Carr *et al.*, 1997; Brookmeyer *et al.*, 2007]. Sin duda, se trata de uno de los retos más importantes a los que deberá enfrentarse la sociedad en el transcurso del siglo XXI.

La enfermedad de Alzheimer se caracteriza por la pérdida de neuronas y de la sinapsis en la corteza cerebral y en ciertas regiones subcorticales. Esta pérdida resulta en una atrofia de las regiones afectadas, incluyendo una degeneración en el lóbulo temporal y parietal y partes de la corteza frontal y la circunvolución cingulada [Wenk, 2003]. El cerebro de las personas que sufren AD presenta dos estructuras típicas: las placas de amiloides y los ovillos neurofibrilares, que sólo son visibles cuando se observan bajo el microscopio [Tiraboschi *et al.*, 2004]. Las placas son depósitos densos e insolubles de la proteína beta-amiloide y de material celular, que se localizan alrededor de las neuronas. Estas placas continúan creciendo hasta formar fibras entreteljadas dentro de la célula nerviosa, los llamados ovillos. Es probable que muchos individuos, en su vejez, desarrollen estas placas y ovillos como parte del proceso normal de envejecimiento, sin embargo, los pacientes con Alzheimer tienen un mayor número en lugares específicos del cerebro como el lóbulo temporal [Bouras *et al.*, 1994].

Por desgracia, actualmente la enfermedad de Alzheimer no tiene cura y la medicación existente sólo consigue mitigar sus síntomas. Sin embargo, en los últimos años se han desarrollado nuevos tratamientos más efectivos que consiguen alargar la vida del paciente y mejorar su calidad de vida.

1.3.2. Diagnóstico de la enfermedad de Alzheimer

Un diagnóstico temprano de la enfermedad permite comenzar antes con el tratamiento, lo que redundará en un alargamiento de la vida del paciente. Además, ayuda a sus familias a planear el futuro. Les da tiempo para discutir las opciones de cuidado mientras el paciente todavía puede participar en la toma de decisiones.

Tradicionalmente, el diagnóstico de AD se lleva a cabo a través de entrevistas con el paciente y mediante la realización de diversos tests para medir sus habilidades mentales. Sin embargo, además de la subjetividad inherente a esta forma de diagnóstico, el hecho de que la enfermedad afecte principalmente a personas de edad avanzada, hace más complejo el diagnóstico precoz. Que una

persona de entre 70 y 80 años tenga pequeños olvidos o dificultades para recordar ciertas cosas puede deberse simplemente al deterioro causado por la edad sin que ello implique la presencia de AD. Debido a esto, el diagnóstico tradicional se limita a clasificar los pacientes como “posible” Alzheimer ó “probable” Alzheimer. Actualmente, la única forma de diagnosticar AD con un 100 % de seguridad es mediante un estudio post-muerte que examine cerebro y detecte la posible presencia de placas de amiloides y/o ovillos neurofibrilares.

La precisión del diagnóstico “en vida” de la enfermedad aumenta sensiblemente si se hace uso de imágenes funcionales del cerebro tales como la obtenidas con técnicas como PET y SPECT. Históricamente, el patrón que se ha asociado a la enfermedad de Alzheimer se ha definido relativo a las áreas corticales, que son más fácilmente identificables en las imágenes funcionales. Basándose en la inspección visual de anormalidades en estas áreas, las imágenes son clasificadas como indicativas de AD cuando se encuentran: *i*) bajos niveles de captación en regiones locales del córtex, en los lóbulos parietal, temporal y/o frontal, o *ii*) un nivel global reducido de captación, relativo al córtex sensoriomotor y visual, tálamo, el ganglio basal y el cerebelo [Hoffman & Phelps, 1986]. Usando estos criterios, el diagnóstico visual de AD obtiene una precisión mayor que los criterios clínicos (77 %) [Silverman *et al.*, 2003, 2001b]. No obstante, el examen visual es complejo y no está exento de dificultades, por lo que la precisión que se obtiene en el diagnóstico precoz efectuado de esta forma, no supera el 80 % [Cummings *et al.*, 1998; Carr *et al.*, 1997].

En la actualidad existe un esfuerzo de investigación importante centrado en el desarrollo de sistemas informáticos que faciliten el procesamiento y uso de imágenes funcionales para el diagnóstico de AD [Stoeckel *et al.*, 2004; Górriz *et al.*, 2008; Ramírez *et al.*, 2008a; I.A.Illan, 2009]. Sin embargo, todavía resulta necesario introducir nuevas técnicas más eficientes de procesado, modelado y clasificación de imágenes, así como mejorar el rendimiento de las técnicas existentes. En el capítulo 9 se muestra como las técnicas de análisis de imagen presentadas en este trabajo suponen un avance significativo en este sentido.

1.4. Estructura de la Tesis Doctoral

Este trabajo está estructurado en cuatro grandes partes. La primera de ellas contiene tres capítulos que tratan sobre imagen médica y su uso para el diagnóstico de enfermedades neurodegenerativas como AD:

- Imagen funcional. Breve disertación sobre las técnicas tomográficas usadas en medicina nuclear, centrándonos en las técnicas de imagen funcional.
-

Asimismo se describe el proceso de adquisición de las imágenes y su posterior reconstrucción y normalización.

- Base de datos de imágenes. Detalle de las bases de datos de imágenes funcionales que se usan en esta tesis con la finalidad de validar los métodos propuestos en la segunda parte.
- La enfermedad de Alzheimer. Descripción de la enfermedad, etiología y su anatomía patológica, así como los métodos tradicionales de diagnóstico.

Los fundamentos teóricos se recogen en la segunda parte, divididos en 2 capítulos, a saber:

- Métodos de clasificación. Conceptualización de aprendizaje estadístico y métodos *kernel*. Además se describen los métodos de clasificación y los sistemas de validación usados en este trabajo.
- Mapas Estadísticos Paramétricos. Descripción de esta herramienta que se ha convertido en un referente para el tratamiento y análisis de imágenes médicas.

La tercera parte abarca cuatro capítulos con las aportaciones de esta tesis:

- Mezcla de Gaussianas. Descripción de un método de extracción de características basado en el modelo de Mezcla de Gaussianas (GMM).
- Mínimos Cuadrados Parciales. Descripción de un algoritmo de reducción de la dimensionalidad basado en la técnica de Mínimos Cuadrados Parciales (PLS).
- Experimentos y resultados. Contiene los experimentos diseñados para comprobar la validez de los métodos descritos en los dos capítulos anteriores.
- Herramienta para estudio y clasificación de imágenes médicas. Presentación de la herramienta informática implementada. Esta herramienta pretende hacer accesible al personal médico los últimos avances realizados en sistemas CAD, incluidos los desarrollados en esta tesis.

Por último, la cuarta parte contiene un amplio resumen en inglés del trabajo presentado.

1.5. Publicaciones científicas

1.5.1. Artículos en revistas internacionales

El trabajo presentado en esta tesis ha derivado en las siguientes publicaciones científicas:

1. **Autores:** F. Segovia, J.M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, M. López y R. Chaves.
Título: Classification of functional brain images using a GMM-based multi-variate approach.
Revista: Neuroscience Letters, Vol. 474 (2010) pp. 58-62
 2. **Autores:** J.M. Górriz, F. Segovia, J. Ramírez, A. Lassi y D. Salas-Gonzalez.
Título: GMM based SPECT image classification for the diagnosis of Alzheimer's disease.
Revista: Applied Soft Computing, Vol. XX (2010) pp. XX-XX
 3. **Autores:** J. Ramírez, J. M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López y I. Álvarez.
Título: Computer Aided Diagnosis System for the Alzheimer's Disease based on Partial Least Squares and Random Forest SPECT image classification.
Revista: Neuroscience Letters, Vol. 472 (2010) pp. 99-103
 4. **Autores:** I.A. Illán, J.M. Górriz, M.M. López, J. Ramírez, D. Salas-Gonzalez, F. Segovia, R. Chaves y C.G. Puntonet.
Título: Computer aided diagnosis of Alzheimer's disease using component based SVM.
Revista: Applied Soft Computing, Vol. XX (2010) pp. XX-XX
 5. **Autores:** D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. Álvarez, M. López, F. Segovia y C. G Puntonet.
Título: Two approaches for selecting set of voxels for the diagnosis of Alzheimer's using brain SPECT images.
Revista: Digital Signal Processing, Vol. XX (2010) pp. XX-XX
 6. **Autores:** M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, M. Gómez-Río y Alzheimer's Disease Neuroimaging Initiative.
-

-
- Título:** Principal Component Analysis-Based Techniques and Supervised Classification Schemes for the Early Detection of the Alzheimer's Disease.
Revista: Neurocomputing, Vol. XX (2010) pp. XX-XX
7. **Autores:** D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, M. López, **F. Segovia**, R. Chaves, y P. Padilla.
Título: Feature selection using factor analysis for Alzheimer's diagnosis using 18F-FDG PET images.
Revista: Medical Physics, Vol. 37 (2010) pp. 6084-6095
 8. **Autores:** P. Padilla, J.M. Górriz, J. Ramírez, E.W. Lang, R. Chaves, **F. Segovia**, M. López, D. Salas-González y I. Álvarez.
Título: Analysis of SPECT Brain Images for the Diagnosis of Alzheimer's Disease Based on NMF for Feature Extraction.
Revista: Neuroscience Letters, Vol. 479 (2010) pp. 192-196
 9. **Autores:** D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, M. López, I. Álvarez, **F. Segovia**, R. Chaves y C.G. Puntonet.
Título: Computer aided diagnosis of Alzheimer's disease using support vector machines and classification trees.
Revista: Physics in Medicine and Biology, Vol. 55 (2010) pp. 2807-2817
 10. **Autores:** P. Padilla, J.M. Górriz, J. Ramírez, R. Chaves, **F. Segovia**, I. Álvarez, D. Salas-González, M. López y C.G. Puntonet.
Título: Alzheimer's disease detection in functional images using 2D Gabor wavelet analysis. Ref. 1
Revista: IET Electronic Letters Journal, Vol. 46 (2010) pp. 556-558
 11. **Autores:** I. A. Illan, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, **F. Segovia** y C. G. Puntonet.
Título: Projecting independent components of SPECT images for computer aided diagnosis of Alzheimer's disease.
Revista: Pattern Recognition Letters, Vol. 46 (2010) pp. 233-238
 12. **Autores:** M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, **F. Segovia** y R. Chaves.
Título: SVM-based CAD System for Early Detection of the Alzheimer's Disease using Kernel PCA and LDA.
Revista: Neuroscience Letters, Vol. 46 (2009) pp. 233-238
 13. **Autores:** R. Chaves, J. Ramírez, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez y **F. Segovia**.
-

Título: SVM-based computer aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting
Revista: Neuroscience Letters, Vol. 461 (2009) pp. 293-297

14. **Autores:** D. Salas-González, J.M. Górriz, J. Ramírez, M. López, I.A. Illan, **F. Segovia**, C.G. Puntonet y M. Gómez-Río.

Título: Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines

Revista: Neuroscience Letters, Vol. 469 (2009) pp. 60-64

15. **Autores:** J. Ramírez, J.M. Górriz, R. Chaves, M. López, D. Salas-Gonzalez, I. Álvarez y **F. Segovia**.

Título: SPECT image classification using random forests

Revista: IET Electronic Letters Journal, Vol. 45 (2009) pp. 604-605

16. **Autores:** I. Álvarez, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C.G. Puntonet y **F. Segovia**.

Título: Alzheimer's Diagnosis Using Eigenbrains and Support Vector Machines

Revista: IET Electronic Letters Journal, Vol. 45 (2009) pp. 342-343

17. **Autores:** M. López, J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, I. Álvarez, **F. Segovia** y C. G. Puntonet

Título: Automatic Tool for the Alzheimer's Disease Diagnosis Using PCA and Bayesian Classification Rules

Revista: IET Electronic Letters Journal, Vol. 45 (2009) pp. 389-391

Además está enviado y pendiente de resolución el siguiente artículo:

- **Autores:** **F. Segovia**, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, I. Alvarez, M. López, R. Chaves, P. Padilla y Alzheimer's Disease Neuroimaging Initiative.

Título: A comparative study of feature extraction methods for the diagnosis of Alzheimer's Disease using the ADNI database.

Revista: Neurocomputing

1.5.2. Aportaciones a congresos internacionales

Por otro lado, el trabajo aquí presentado ha dado lugar a las siguientes aportaciones a congresos:

-
1. **Autores:** F. Segovia, J. M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, M. López, G. C. Puntonet.
Título: Tool for Study and Classification of Medical Images.
Tipo de participación: Poster.
Congreso: The fifth IADAT International Conference on Education IADATe-2009.
Lugar celebración: Bilbao (Spain).
Fecha: Junio 2009
 2. **Autores:** F. Segovia, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, I. A. Illán, M. López, R. Chaves, C. G. Puntonet.
Título: Automatic selection of ROIs using a model-based clustering approach.
Tipo de participación: Poster.
Congreso: Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC 2009)
Lugar celebración: Orlando (USA)
Fecha: Octubre 2009
 3. **Autores:** F. Segovia, J. Ramírez, J. M. Górriz, R. Chaves, D. Salas-Gonzalez, M. López y I. Alvarez y P. Padilla y C. G. Puntonet.
Título: Partial Least Squares for feature extraction of SPECT images.
Tipo de participación: Presentación oral.
Congreso: The 5th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010)
Lugar celebración: San Sebastian (Spain)
Fecha: Junio 2010
 4. **Autores:** J. Ramírez, J.M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, I. Álvarez y P. Padilla.
Título: Early Alzheimer's disease diagnosis using Partial Least Squares and Random Forests.
Tipo de participación: Poster.
Congreso: The IEEE International Symposium on Biomedical Imaging (ISBI 2010)
Lugar celebración: Rotterdam (Netherlands)
Fecha: Abril 2010
 5. **Autores:** D. Salas-Gonzalez, J. Estrada, J. M. Górriz, J. Ramirez, F. Segovia, R. Chaves, M. Lopez y I. A. Illán.
-

Título: Improving the Convergence Rate in Affine Registration of PET Brain Images Using Histogram Matching.

Tipo de participación: Poster.

Congreso: Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC 2010)

Lugar celebración: Knoxville, Tennessee (USA)

Fecha: Noviembre 2010

6. **Autores:** I. A. Illán, J. M. Górriz, J. Ramírez, R. Chavez, **F. Segovia**, M. López, D. Salas-Gonzalez y C. G. Puntonet.

Título: Machine Learning for Very Early Alzheimer's Disease Diagnosis; a 18F-FDG and PiB PET Comparison.

Tipo de participación: Poster.

Congreso: Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC 2010)

Lugar celebración: Knoxville, Tennessee (USA)

Fecha: Noviembre 2010

7. **Autores:** R. Chaves, J. Ramírez, J.M. Górriz, M. López, D. Salas, I. Illán, **F. Segovia** y P. Padilla.

Título: Effective Diagnosis of Alzheimer's Disease by means of Association Rules.

Tipo de participación: Presentación oral.

Congreso: The 5th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010)

Lugar celebración: San Sebastian (Spain)

Fecha: Junio 2010

8. **Autores:** I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, P. Padilla, R. Chaves, **F. Segovia** y C.G. Puntonet.

Título: Exploring Symmetry to Assist Alzheimer's Disease Diagnosis.

Tipo de participación: Presentación oral.

Congreso: The 5th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010)

Lugar celebración: San Sebastian (Spain)

Fecha: Junio 2010

9. **Autores:** D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, **F. Segovia**, R. Chaves, M. López y I.A. Illán.

Título: Multivariate Approaches for Alzheimer's Disease Diagnosis Using

Bayesian Classifiers.

Tipo de participación: Presentación oral.

Congreso: The 5th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010)

Lugar celebración: San Sebastian (Spain)

Fecha: Junio 2010

10. **Autores:** D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, I. Álvarez, M. López, **F. Segovia** y M. Gómez-Río.

Título: Skewness as feature for the diagnosis of Alzheimer's disease using SPECT images.

Tipo de participación: Poster.

Congreso: IEEE International Conference on Image Processing (ICIP 2009)

Lugar celebración: El Cario (Egypt)

Fecha: Noviembre 2009

11. **Autores:** M. López, J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, I. Álvarez, **F. Segovia**, R. Chaves.

Título: Multivariate Approaches for Alzheimer's Disease Diagnosis Using Bayesian Classifiers.

Tipo de participación: Poster.

Congreso: Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC 2009)

Lugar celebración: Orlando (USA)

Fecha: Octubre 2009

12. **Autores:** M. López, J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, I. Álvarez, **F. Segovia**, R. Chaves.

Título: Neurological Image Classification for the Alzheimer's Disease Diagnosis Using Kernel PCA and Support Vector Machines.

Tipo de participación: Poster.

Congreso: Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC 2009)

Lugar celebración: Orlando (USA)

Fecha: Octubre 2009

13. **Autores:** R. Chaves, J. Ramírez, J. M. Górriz, M. López, I. Álvarez, D. Salas-Gonzalez, **F. Segovia**.

Título: SPECT Image Classification Based on NMSE Feature Correlation

Weighting and SVM.

Tipo de participación: Poster.

Congreso: Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC 2009)

Lugar celebración: Orlando (USA)

Fecha: Octubre 2009

14. **Autores:** J. Ramírez, R. Chaves, J. M. Górriz, M. López, I. Álvarez, D. Salas-Gonzalez, **F. Segovia**.
Título: Computer Aided Diagnosis of the Alzheimer's Disease Combining SPECT-Based Feature Selection and Random Forest Classifiers.
Tipo de participación: Poster.
Congreso: Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC 2009)
Lugar celebración: Orlando (USA)
Fecha: Octubre 2009
 15. **Autores:** J.M. Górriz, J. Ramírez, A. Lassl, I. Álvarez, **F. Segovia**, D. Salas, and M. López
Título: Classification of SPECT Images Using Clustering Techniques Revisited.
Tipo de participación: Poster.
Congreso: International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC 2009)
Lugar celebración: Santiago de Compostela (Spain).
Fecha: Junio 2009
 16. **Autores:** D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. Álvarez, **F. Segovia**, C. G. Puntonet.
Título: Analysis of Brain SPECT Images for the diagnosis of Alzheimer disease using first and second order moments.
Tipo de participación: Poster.
Congreso: International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC 2009)
Lugar celebración: Santiago de Compostela (Spain).
Fecha: Junio 2009
 17. **Autores:** J. Ramírez, R. Chaves, J. M. Górriz, I. Álvarez, M. López, D. Salas-Gonzalez, **F. Segovia**.
Título: Functional Brain Image Classification Techniques for Early Alzheimer Disease Diagnosis.
-

-
- Tipo de participación:** Poster.
Congreso: International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC 2009)
Lugar celebración: Santiago de Compostela (Spain).
Fecha: Junio 2009
18. **Autores:** M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, **F. Segovia**, M. Gómez-Río.
Título: Support Vector Machines and Neural Networks for the Alzheimer's Disease Diagnosis Using PCA.
Tipo de participación: Poster.
Congreso: International Work-conference on the Interplay between Natural and Artificial Computation (IWINAC 2009)
Lugar celebración: Santiago de Compostela (Spain).
Fecha: Junio 2009
19. **Autores:** D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. Álvarez, **F. Segovia**, C. G. Puntonet.
Título: Selecting regions of interest for the diagnosis of Alzheimer's disease in brain SPECT images using Welch's t-test.
Tipo de participación: Poster.
Congreso: International Work-Conference on Artificial Neural Networks (IWANN 2009)
Lugar celebración: Salamanca (Spain).
Fecha: Junio 2009
20. **Autores:** M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, **F. Segovia** and C.G. Puntonet.
Título: Automatic System for Alzheimer's Disease Diagnosis Using Eigenbrains and Bayesian Classification Rules.
Tipo de participación: Poster.
Congreso: International Work-Conference on Artificial Neural Networks (IWANN 2009)
Lugar celebración: Salamanca (Spain).
Fecha: Junio 2009
21. **Autores:** J. Ramírez, R. Chaves, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, **F. Segovia**.
Título: SPECT image classification techniques for computer aided diagnosis of the Alzheimer disease.
-

Tipo de participación: Poster.

Congreso: International Work-Conference on Artificial Neural Networks (IWANN 2009)

Lugar celebración: Salamanca (Spain).

Fecha: Junio 2009

22. **Autores:** I. Álvarez, J. M. Górriz, J. Ramírez, M. López, D. Salas-Gonzalez, C. G. Puntonet, **F. Segovia**, B. Prieto.

Título: Alzheimer's Diagnosis Using Eigenbrains and Support Vector Machines.

Tipo de participación: Poster.

Congreso: International Work-Conference on Artificial Neural Networks (IWANN 2009)

Lugar celebración: Salamanca (Spain).

Fecha: Junio 2009

23. **Autores:** M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, **F. Segovia** and C. G. Puntonet

Título: Computer Aided Diagnosis of Alzheimer Disease Using Principal Components and Bayesian Classifiers

Tipo de participación: Poster

Congreso: International Symposium on Neural Networks (ISNN 2009)

Lugar celebración: Wuhan (China)

Fecha: Mayo 2009

24. **Autores:** D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. Álvarez, M. López, **F. Segovia** and C. G. Puntonet

Título: Selecting regions of interest for the diagnosis of Alzheimer using brain SPECT images

Tipo de participación: Poster

Congreso: International Symposium on Neural Networks (ISNN 2009)

Lugar celebración: Wuhan (China)

Fecha: Mayo 2009

25. **Autores:** I. Álvarez, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C. G. Puntonet and **F. Segovia**

Título: Independent Component Analysis of SPECT Images to assist the Alzheimer's Disease Diagnosis

Tipo de participación: Poster

Congreso: International Symposium on Neural Networks (ISNN 2009)

-
- Lugar celebración:** Wuhan (China)
Fecha: Mayo 2009
26. **Autores:** J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, M. López, **F. Segovia** and C. G. Puntonet
Título: Effective detection of the Alzheimer disease by means of coronal NMSE SVM feature classification
Tipo de participación: Poster
Congreso: International Symposium on Neural Networks (ISNN 2009)
Lugar celebración: Wuhan (China)
Fecha: Mayo 2009
27. **Autores:** D. Salas-González, J. M. Górriz, J. Ramírez, M. López, I. Álvarez, **F. Segovia** and C. G. Puntonet
Título: Computer Aided Diagnosis of Alzheimer Disease Using Support Vector Machines and Classification Trees
Tipo de participación: Poster
Congreso: International Conference on Neural Information Processing (ICONIP 2008)
Lugar celebración: Auckland (Nueva Zelanda)
Fecha: Noviembre 2008
28. **Autores:** J. Ramírez, J. M. Górriz, M. López, D. Salas-González, I. Álvarez, **F. Segovia** and C. G. Puntonet
Título: Early detection of the Alzheimer disease combining feature selection and kernel machines
Tipo de participación: Poster
Congreso: International Conference on Neural Information Processing (ICONIP 2008)
Lugar celebración: Auckland (Nueva Zelanda)
Fecha: Noviembre 2008
29. **Autores:** I. Álvarez, M. López, J.M. Górriz, J. Ramírez, C.G. Puntonet, D. Salas, **F. Segovia**
Título: Automatic Classification System for the Diagnosis of Alzheimer Disease Using Component-Based SVM Aggregations
Tipo de participación: Poster
Congreso: International Conference on Neural Information Processing (ICONIP 2008)
Lugar celebración: Auckland (Nueva Zelanda)
Fecha: Noviembre 2008
-

1.5.3. Capítulos de libros

1. **Autores:** F. Segovia, J.M. Górriz, J.Ramírez, D. Salas-Gonzalez, I.A. Illán, M. López, R. Chaves y P. Padilla
Título: A Comparison of Univariate and Multivariate Supervised Learning Approaches for Classification of SPECT Images
Libro: Recent Advances in Biomedical Signal Processing, Bentham Science Publishers, 2010
 2. **Autores:** I. A. Illán, M. López, J. M. Górriz, J. Ramírez, F. Segovia, D. Salas-Gonzalez, R. Chaves y C.G. Puntonet
Título: Functional Image Classification Techniques for Early Alzheimer's Disease Detection
Libro: Recent Advances in Biomedical Signal Processing, Bentham Science Publishers, 2010
 3. **Autores:** R. Chaves, D. Salas-Gonzalez, J. Ramirez, J.M. Gorriz, M. López, I. Álvarez y F. Segovia
Título: Preprocessing of Functional Brain Imaging for Computer Aided Diagnosis Systems
Libro: Recent Advances in Biomedical Signal Processing, Bentham Science Publishers, 2010
-

Parte I

Técnicas de imagen médica usadas en el diagnóstico de enfermedades neurodegenerativas

CAPÍTULO

2

TÉCNICAS TOMOGRÁFICAS USADAS EN MEDICINA NUCLEAR

En este capítulo se describirán las principales técnicas de imagen usadas en medicina nuclear, centrándonos en las técnicas de imagen funcional por su mayor utilidad en el diagnóstico de enfermedades neurológicas. A diferencia de otros tipos de imágenes médicas más comunes, las imágenes funcionales muestran la actividad fisiológica dentro de un determinado tejido o en un órgano. Se obtienen mediante el empleo de sustancias conocidas como radiofármacos que, una vez en el interior del organismo, se descomponen emitiendo partículas que son detectables desde el exterior, y que permiten saber como se distribuye el radiofármaco dentro del cuerpo. En la última sección de este capítulo se detalla el preprocesado que deben seguir las imágenes funcionales antes de ser usadas en un análisis.

2.1. Introducción

Son varias las modalidades de imagen médica que se usan en medicina nuclear para el diagnóstico y seguimiento de diversas demencias, si bien, podemos dividir las en dos ramas. Por un lado encontramos aquellas que proporcionan información sobre las estructuras internas del organismo, como la Resonancia Magnética (MRI, del inglés Magnetic Resonance Imaging) o Tomografía Axial Computarizada (TAC ó, en inglés, CT). Y, por otro lado, tenemos las técnicas que proporcionan imágenes con información sobre el funcionamiento del organismo como son la Tomografía Computarizada por Emisión de un Solo Fotón (SPECT) y la Tomografía por Emisión de Positrones (PET).

Todas estas técnicas presentan dos características que han resultado fundamentales para su expansión, hasta convertirse en un procedimiento estandarizado para diagnóstico de diversas patologías en miles de hospitales y centros médicos de todo el mundo:

- Son técnicas no invasivas, por lo que apenas causan molestias a los pacientes y es posible aplicarlas de forma rutinaria.
- Proporcionan una visión amplia de la zona bajo estudio en contraposición a otras modalidades de imagen que se centran en determinados tejidos o grupos celulares. Una sola imagen recoge toda la información sobre la actividad de una zona u órgano concreto.

2.2. Técnicas de neuroimagen estructural

Uno de los objetivos principales del uso de técnicas de imagen estructural, como MRI o CT, para el diagnóstico de enfermedades neurodegenerativas es la exclusión de causas tratables o reversibles de demencia. Un gran número de informes avisan de que hasta un 20% de los pacientes con síntomas de demencia presentan anomalías cerebrales susceptibles de tratamiento quirúrgico, como hidrocefalia de presión normal, hematoma subdural y tumores. Cada una de estas patologías deja marcas típicas en las imágenes estructurales por lo que, tanto CT como MRI convencional, a menudo permiten un diagnóstico definitivo.

Las imágenes estructurales tienen un papel menos importante en el diagnóstico de otras formas de demencia, tales como las relacionadas con la encefalopatía inflamatoria, tóxica o metabólica. Sin embargo, MRI y CT pueden ayudar al diagnóstico diferencial, ya que algunas de estas formas de demencia se asocian

con hallazgos radiológicos típicos. Las lesiones del cuerpo calloso y las lesiones simétricas de la corona radiada y el centro semioval son visibles en imágenes MRI y se asocian con el síndrome de Marchiafava-Bignami [Ferracci *et al.*, 1999]. Incrementos bilaterales de la intensidad de la señal en los ganglios basales, y estructuras periventriculares demasiado planas y rayadas, son síntomas de la enfermedad de Creutzfeldt-Jacob [Krüger *et al.*, 1990; Milton *et al.*, 1991]. Las imágenes cerebrales estructurales tiene un papel bien definido en la diferenciación de las formas degenerativas de la demencia vascular y en la evaluación de la gravedad de algunos cambios estructurales, como la atrofia cortical o leucoaraiosis, asociados con la demencia. Su papel en la distinción entre los cambios asociados con el envejecimiento y las personas con demencia, y en la correlación de anomalías estructurales del cerebro con deterioro clínico es cada vez mayor. En el diagnóstico de AD, la disminución del volumen cerebral global se relaciona con la tasa de cambio en las puntuaciones del examen MMSE. La evaluación cuantitativa de la formación del hipocampo por resonancia magnética volumétrica es una valiosa herramienta en la evaluación de pacientes con AD. En pacientes sanos de avanzada edad se ha observado una disminución anual significativa en el volumen del hipocampo y un aumento en el volumen del cuerno temporal. Estas tasas fueron de aproximadamente 2,5 veces mayores en los pacientes con AD. Del mismo modo, otros estudios de resonancia magnética volumétrica mostraron que el volumen del hipocampo era significativamente menor en pacientes con AD que en controles, con una relación significativa entre los volúmenes del hipocampo, la severidad de la demencia y la edad del paciente.

2.3. Técnicas de neuroimagen funcional

Las técnicas de imagen funcional (IF), tales como SPECT y PET juegan un papel consolidado en la investigación de las enfermedades neurodegenerativas [Herholz, 1995; Scheltens, 1999; Iacoboni *et al.*, 1999]. Ambas técnicas proporcionan imágenes que representan la distribución en el cerebro de un proceso bioquímico específico. Para ello hacen uso de sustancias radioactivas conocidas como radiofármacos que son ingeridas por el paciente o se inyectan vía intravenosa. Cuando el radiofármaco se desintegra en el interior del organismo emite ciertas partículas que son detectables desde el exterior (con la ayuda de cámaras al efecto) y que permiten conocer la distribución del radiofármaco y, a partir de ella, inferir el funcionamiento de ciertos órganos. Alteraciones como la disminución de la actividad metabólica o la reducción del flujo sanguíneo se reflejan fielmente en las imágenes funcionales. La reducción metabólica observada con

la medición del consumo de oxígeno y glucosa, acompañada de una reducción de la perfusión regional, no puede ser explicada simplemente como consecuencia de la pérdida de tejido [Ibanez *et al.*, 1998].

En la actualidad, las mediciones del flujo sanguíneo regional cerebral con SPECT y mediciones regionales de la glucosa o el flujo sanguíneo cerebral con PET se utilizan para los estudios clínicos de varios trastornos degenerativos, como AD. Hay evidencias de un patrón de hipoperfusión bilateral en las regiones temporoparietales desde la fase temprana de la enfermedad. En estos casos, la reducción del metabolismo es mucho mayor que la disminución *fisiológica* observada en el envejecimiento normal, y tiene una distribución espacial típica, con la participación de la corteza asociativa en las áreas escamosas de ambos hemisferios [Herholz, 1995]. Se ha demostrado que un examen neuropsicológico junto con la evaluación de imágenes funcionales del cerebro tomadas con las técnicas de PET o SPECT pueden mejorar el diagnóstico de AD.

El rápido avance de las técnicas de imagen funcional ha sido posible gracias a dos avances científicos. Por un lado, la aparición de radiofármacos seguros que puedan ser administrados a seres humanos así como el desarrollo de cámaras capaces de detectar de forma precisa la radiación emitida durante la descomposición de los radiofármacos. Y, por otro lado, el desarrollo de sistemas informáticos que permiten la reconstrucción de la imagen, a partir de la información proporcionada por los detectores, en un tiempo aceptable.

2.3.1. SPECT

La tomografía por emisión de un solo fotón (SPECT, del inglés Single Photon Emission Computed Tomography) [English & Childs, 1996; Ayache, 1996] es una técnica tomográfica de medicina nuclear que se caracteriza por el uso de radioisótopos que son inyectados directamente en el flujo sanguíneo del paciente y que durante su descomposición radioactiva emiten rayos gamma. El radioisótopo usado depende del protocolo de adquisición, si bien, en la mayoría de los escáneres cerebrales se suele usar el Tecnecio 99^m en combinación con los radiofármacos HMPAO ó ECD que dan lugar a los trazadores 99^mTc-HMPAO y 99^mTc-ECD . La tabla 2.1 muestra una comparativa con las principales características de ambos trazadores.

Las imágenes SPECT miden la distribución del flujo sanguíneo cerebral (rCBF), la cual está relacionado con la actividad del cerebro. Por este motivo, esta modalidad de imagen se suele usar en el diagnóstico de varias demencias entre las que se encuentra la enfermedad de Alzheimer. En la detección de enfermedades neurodegenerativas, una disminución del flujo sanguíneo en la región

Tabla 2.1: Principales características de los dos trazadores usados para la obtención de imágenes SPECT cerebrales.

	^{99m}Tc -HMPAO	^{99m}Tc -ECD
Vida media (hora)	72	24
Eficiencia de extracción al primer paso (%)	80	70
Redistribución cerebral	No	No
Actividad en sangre en 1 hora (%)	12	4.9
Estabilidad tras el marcaje (horas)	0.5	más de 6
Contraste de imagen	Menor	Mayor
Dosis efectiva cuerpo entero (mSr)	5-6.2	4.3

frontotemporal se considera una prueba de la presencia de demencia frontotemporal [Miller *et al.*, 1991; Starkstein *et al.*, 1994]. Asimismo, una anormalidad en el flujo sanguíneo cerebral en la zona posterior es un claro síntoma de Alzheimer [Neary *et al.*, 1987; Burns *et al.*, 1989; McKeith *et al.*, 1993; Montaldi *et al.*, 1990]. La figura 2.1 muestra las diferencias en el rCBF de una persona sana y otra que sufre Alzheimer en un estado avanzado de la enfermedad.

2.3.2. PET

La tomografía por emisión de positrones (PET, del inglés Positron Emission Tomography) es una técnica de medicina nuclear que produce imágenes funcionales de procesos del interior del cuerpo. La característica principal de esta modalidad de imagen es el uso de isótopos radioactivos que al descomponerse emiten positrones. Estas partículas se aniquilan al chocar con los electrones del interior del cuerpo generando dos rayos gamma que son emitidos en direcciones opuestas. Por lo tanto, al igual que en SPECT, se emplea un detector de rayos gamma para capturar la imagen, si bien, en este caso contamos con dos rayos gamma por cada positrón. Esto permite determinar de forma más precisa la localización de las partículas radioactivas, lo que dota a las imágenes PET de mayor resolución que las SPECT.

En función del radiofármaco usado es posible estudiar distintos procesos biológicos. Por ejemplo, el isótopo ^{15}O se emplea en combinación con H_2O como trazador para estudiar la perfusión cerebral. No obstante, la principal aplicación clínica de PET consiste en el estudio del metabolismo y, en este caso, el isótopo usado es el Fluor 18 que se combina con un análogo de la glucosa dando lugar a

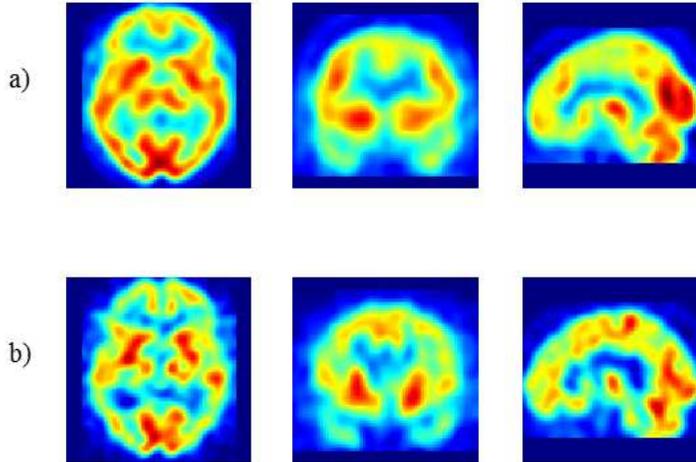


Figura 2.1: Diferentes cortes de dos imágenes SPECT. La fila superior muestra tres cortes de una imagen SPECT cerebral correspondiente a una persona sana. Los tres cortes de la fila inferior corresponden a la imagen SPECT de una persona que sufre Alzheimer en un estado avanzado de la enfermedad

la fluorodesoxiglucosa (^{18}F -FDG). Una vez introducida en el flujo sanguíneo, la ^{18}F -FDG es almacenada por las células pero a diferencia de la glucosa normal, no puede utilizarla para producir energía por lo que queda atrapada dentro de ellas. De este modo podemos medir la captación de glucosa por parte de las células que es indicativa de la actividad de las mismas.

La disminución de la actividad cerebral en determinadas zonas es un síntoma claro Alzheimer. Por ello las imágenes PET son usadas con frecuencia en el diagnóstico de esta enfermedad. Existen abundantes datos que apoyan la afirmación de que las imágenes PET ofrecen una sensibilidad y especificidad bastante altas en la diferenciación entre AD y envejecimiento normal, incluso en etapas tempranas de la enfermedad. Estudios multicéntricos han demostrado que mediante imágenes ^{18}F -FDG PET es posible discriminar entre pacientes con AD probable y controles. Además, estas imágenes se emplean con frecuencia en diagnóstico, tratamiento y evaluación de la respuesta al tratamiento de pacientes con cáncer [Delbeke *et al.*, 1999; Eary, 1999].

Una de las principales desventajas de PET frente a otras modalidades de

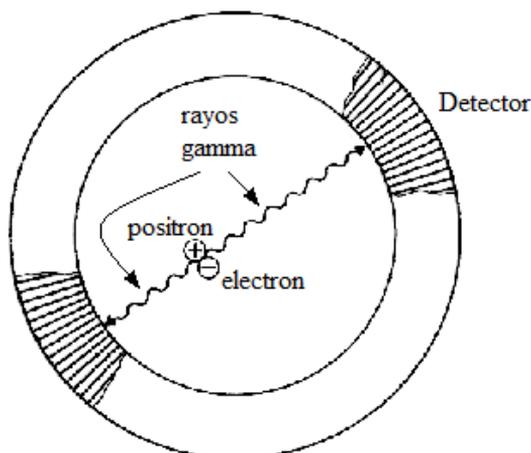


Figura 2.2: Principio físico de PET

imagen es su precio. De hecho, el elevado coste de los radiofármacos usados debido a la corta vida media de los mismos, es el motivo principal que hace que esta técnica esté menos extendida que SPECT (que aporta imágenes con una menor resolución pero a menor precio).

2.3.3. Adquisición de las imágenes

La adquisición de las imágenes, tanto SPECT como PET, se realiza con el paciente en posición decúbito supino (acostado boca arriba, cuello en posición neutra, mirada dirigida al cént, miembros superiores extendidos y pegados al tronco, palmas de las manos hacia arriba, extremidades inferiores también extendidas, con los pies en flexión neutra y la punta de los dedos gordos hacia arriba) y en una sala sin ruidos y con luz tenue en la que el paciente debe estar al menos 10 minutos antes del comienzo del proceso. Hemos de tener en cuenta que las imágenes que se van a obtener mostrarán el funcionamiento del cerebro y por tanto es conveniente que no haya influencias del ambiente que pudieran provocar una actividad cerebral mayor de lo normal (en estado de reposo). La cabeza del paciente debe estar inmovilizada y el detector debe posicionarse tan próximo al cerebro como sea posible, preferiblemente con un radio de rotación de 14 centímetros o menos desde la superficie del detector de colisiones al centro del cerebro del paciente.

El proceso de adquisición puede comenzar a los 15 minutos de la administración del radiofármaco aunque es recomendable esperar entre 60 y 90 minutos. Durante el proceso, la cámara (o detector) va girando alrededor del cerebro y durante su movimiento va tomando imágenes en 2 dimensiones conocidas como proyecciones. Típicamente se toma una proyección cada 2 grados de giro, lo que hace un total de 180 proyecciones para cada imagen cerebral. Finalmente, la imagen tridimensional del cerebro se reconstruye, con ayuda de un ordenador, a partir de las proyecciones usando el algoritmo de retroproyección filtrada (FBP) descrito a continuación, en combinación con un filtro Butterworth para la eliminación del ruido.

2.3.4. Reconstrucción de las imágenes

La reconstrucción de una imagen funcional es el proceso a partir del cual se genera un mapa de activación tridimensional a partir de las proyecciones tomadas por la cámara de rayos gamma [Lange & Carson, 1984; Vardi *et al.*, 1985; Hudson & Larkin, 1994; Bruyant, 2002; Chornoboy *et al.*, 1990; Chaves *et al.*, 2010]. Para este proceso se suele emplear el algoritmo de retroproyección filtrada (FBP) que está basado en el teorema de cortes de Fourier.

Teorema de cortes de Fourier. La transformada de Fourier $S_\theta(w)$ de la proyección paralela $P_\theta(t)$ de una imagen $f(x, y)$ tomada con un ángulo θ y definida del siguiente modo:

$$S_\theta(w) = \int_{-\infty}^{+\infty} P_\theta(t) \exp(-j2\pi wt) dt, \quad (2.1)$$

proporciona un corte de la transformada de Fourier bidimensional:

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \exp(-j2\pi(ux + vy)) dx dy, \quad (2.2)$$

que se encuentra a un ángulo θ del eje u , es decir,

$$S_\theta(w) = F(u = w \cos \theta, v = w \sin \theta). \quad (2.3)$$

El teorema de cortes de Fourier relaciona las medidas de los datos de proyección con la transformada de Fourier de la sección eficaz del objeto. La utilidad del teorema radica en que tomando las proyecciones de un objeto a ángulos $\theta_1, \theta_2, \dots, \theta_k$ y tomando la transformada de Fourier de las mismas, podemos

determinar los valores de la transformada bidimensional $F(u, v)$ en líneas radiales del plano uv (dominio de frecuencias espaciales). Si pudiéramos tomar un número infinito de proyecciones entonces conoceríamos $F(u, v)$ en todo el plano y tomando su transformada inversa encontraríamos la imagen $f(x, y)$ del objeto. La transformada inversa viene dada por:

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{2\pi j(ux+uv)} dudv \quad (2.4)$$

En la práctica, la función $F(u, v)$ se conoce sólo a lo largo de un número finito de líneas radiales de modo que se deben interpolar las que falten. Sin embargo, para frecuencias altas, las líneas radiales obtenidas experimentalmente se encuentran muy separadas entre sí, por lo que la interpolación puede inducir un error considerable. Para evitar este problema se emplea un mecanismo conocido como *filtrado* que consiste en una ponderación de cada proyección en el dominio de las frecuencias. Después del *filtrado* se pueden reconstruir los puntos en el dominio espacial correspondientes a cada una de las líneas radiales en el dominio de las frecuencias. Este mecanismo se conoce como *retroproyección*. Por tanto, el algoritmo FBP tiene dos etapas, una del filtrado y otra de retroproyección.

De acuerdo al algoritmo FBP una imagen de la sección cruzada $f(x, y)$ de un objeto se puede recuperar por (etapa de retroproyección):

$$f(x, y) = \int_0^{\pi} Q_{\theta}(x \cos \theta + y \sin \theta) d\theta, \quad (2.5)$$

donde $Q_{\theta}(t)$ representa el filtrado de las proyecciones y se define como:

$$Q_{\theta}(t) = \int_{-\infty}^{+\infty} S_{\theta}(w) |w| \exp(j2\pi wt) dw. \quad (2.6)$$

siendo $S_{\theta}(w)$ la transformada de Fourier de una proyección tal y como se define en el teorema de cortes de Fourier.

Un gran inconveniente de FBP es que se amplifica de forma indeseada el ruido de altas frecuencias impactando sobre la calidad de la imagen. Estos efectos se producen por la de multiplicación de $S_{\theta}(w)$ por $|w|$ en la ecuación 2.6. Para atenuar el ruido de altas frecuencias amplificado durante la reconstrucción FBP se han propuesto diversas funciones tipo ventana. De este modo, el método de reconstrucción descrito por las ecuaciones 2.5 y 2.6 se redefine normalmente aplicando una ventana en frecuencia con valores cercanos a cero cuando la frecuencia tiende a π . Entre las ventanas más comunes para la reconstrucción FBP

se encuentran: sinc (filtro de Shepp-Logan), coseno, Hamming y Hanning. Sin embargo, incluso cuando el ruido de reconstrucción se mantiene bajo usando la aproximación de FBP con control de ruido, se necesita filtrar el ruido capturado por el sistema de adquisición para mejorar la calidad de las imágenes reconstruidas. La mayoría de sistemas de procesamiento de imágenes incluyen un paso de filtrado antes de la reconstrucción para eliminar el ruido captado por la cámara gamma y otro tras la reconstrucción para eliminar el ruido amplificado por el algoritmo FBP.

2.4. Preprocesamiento de las imágenes

Antes de ser utilizadas en un análisis estadístico, las imágenes tomográficas deben seguir un proceso conocido como preprocesamiento que garantiza que las imágenes son comparables entre sí. Este proceso debe incluir al menos la normalización tanto espacial como en intensidad, si bien, en ocasiones acarrea más tareas.

2.4.1. Normalización espacial

La complejidad de las estructuras cerebrales y las diferencias entre cerebros de diferentes pacientes hace necesaria la normalización de las imágenes con respecto a una plantilla común. Este paso nos permite comparar vóxel a vóxel las intensidades de las imágenes cerebrales de pacientes diferentes.

El método de normalización asume un modelo genérico afín con 12 parámetros [Woods, 2000] y una función de coste que presenta un valor extremo cuando la plantilla y la imagen coinciden anatómicamente. La función objetivo que se optimiza es normalmente la diferencia cuadrática media entre ambas imágenes, la imagen fuente y la plantilla de referencia:

$$CF = \sum_i (f(\mathbf{M}\mathbf{x}_i) - g(\mathbf{x}_i))^2, \quad (2.7)$$

donde f denota la imagen original y g la plantilla. Para cada vóxel $\mathbf{x} = (x_1, x_2, x_3)$ de una imagen, la transformación afín a las coordenadas $\mathbf{y} = (y_1, y_2, y_3)$ se ex-

presa mediante la multiplicación matricial $\mathbf{y} = \mathbf{M}\mathbf{x}$.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix}. \quad (2.8)$$

Tras la normalización afín, la imagen resultante se registra usando un modelo de transformación no rígido más complejo. Las deformaciones no lineales se ajustan a una plantilla y se parametrizan mediante una combinación lineal de componentes de frecuencias más bajas de las bases de la transformada coseno tridimensional [Ashburner & Friston, 1999].

La figura 2.3 muestra un ejemplo de la operación del proceso de normalización en imágenes SPECT. La columna izquierda muestra una imagen sin normalizar, la columna central muestra la plantilla usada en el proceso, y finalmente la imagen normalizada correspondiente se muestra en la columna derecha. En dicha figura, para la obtención de la imagen transformada se ha aplicado un modelo de deformación pequeña, y una regularización de la energía de curvatura del campo de desplazamiento. Se observa claramente cómo la imagen transformada se ajusta a la forma de la plantilla. Además ambas tienen el mismo número de vóxeles.

El reto de la normalización espacial supone conseguir minimizar los efectos que producen las diferencias en la imagen debidas a características individuales de cada sujeto sin alterar las diferencias debidas a efectos de la enfermedad, lo que en la práctica es difícil de conseguir completamente. La dificultad aumenta en gran medida debido al hecho de tratar con información funcional, que hace que los límites espaciales de las diferentes regiones cerebrales sean difusos. Los resultados de los análisis posteriores dependerán en gran medida de la normalización espacial.

2.4.2. Normalización de intensidad

Después de la normalización espacial, se requiere una normalización en intensidad para poder comparar entre imágenes de diferentes sujetos. La comparación directa entre la información de la intensidad de los vóxeles, incluso entre diferentes adquisiciones de un mismo sujeto, no es posible sin esta normalización de las intensidades.

Llevando a cabo un proceso parecido a [Saxena *et al.*, 1998], si el nivel de intensidad se normaliza a la intensidad máxima (la cual se calcula individualmente

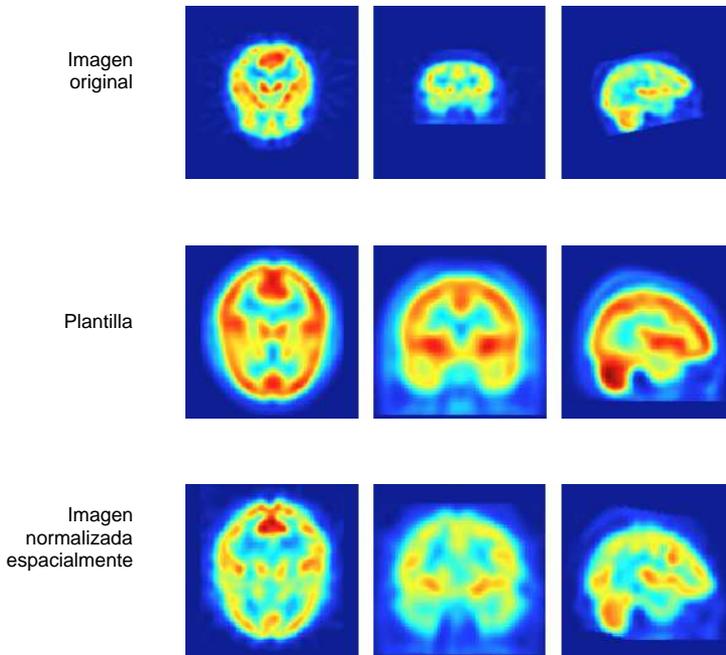


Figura 2.3: Tres imágenes SPECT. *Columna izquierda:* Imagen original. *Columna central:* Plantilla. *Columna derecha:* Imagen transformada tras el proceso de normalización espacial.

para cada volumen calculando la media del 3% de los vóxeles de mayor intensidad) las imágenes pueden estar expuestas a saturación. Como se puede observar en la figura 2.4, los patrones de intensidad que presentan las imágenes pueden ser muy variados, aunque todos presentan una acumulación de intensidad en los primeros niveles de intensidad. Esta acumulación de intensidad se corresponde con las regiones de fuera del cerebro, que suponen una gran cantidad de información irrelevante. Si se toma el 3% de todos los vóxeles, se tendrá en cuenta mucha de esta información lo que producirá que se normalice a un máximo de

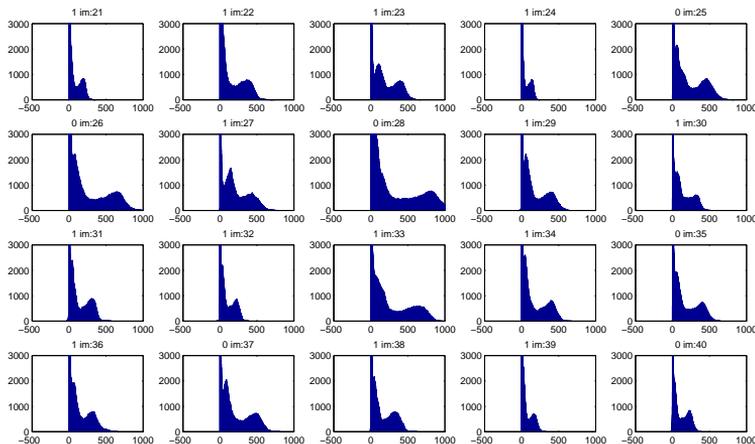


Figura 2.4: 20 histogramas de intensidad de diferentes imágenes SPECT [Salas-Gonzalez, 2009]

intensidad relativamente bajo, y consecuentemente las imágenes saturarán.

Otra propuesta consiste en normalizar la intensidad de las imágenes a un valor máximo I_{max} que se obtiene promediando el 0.1% de los vóxeles de mayor intensidad que superan un umbral. El umbral se fija ajustándolo manualmente a un valor óptimo, obtenido como el valor de intensidad de la décima barra en un histograma de intensidad de 50 barras. De esta forma se asegura que la información irrelevante de baja intensidad contenida fuera del cerebro sea desechada y se previene la saturación.

CAPÍTULO

3

BASES DE DATOS DE IMÁGENES

Los métodos desarrollados en esta tesis, de carácter general, han sido validados mediante tres bases de datos de imágenes tomográficas para la enfermedad de Alzheimer. Entendemos por base de datos un conjunto de imágenes tomográficas cerebrales etiquetadas. Las etiquetas que han sido asignadas a las imágenes permiten distinguir entre imágenes de enfermos de AD (en distintas etapas de la enfermedad), imágenes de personas con deterioro cognitivo leve que podría derivar en Alzheimer (puede considerarse una etapa precoz de la enfermedad) e imágenes de personas que no padecen la enfermedad.

3.1. Base de datos del Hospital Virgen de las Nieves

La primera base de datos usada en este trabajo para evaluar los algoritmos de análisis de imágenes funcionales propuestos, está compuesta por 97 imágenes SPECT que fueron recogidas durante un estudio sobre la enfermedad de Alzheimer llevado a cabo por el Servicio de Medicina Nuclear del Hospital “Virgen de las Nieves” (HVN), situado en Granada (España).

El radiofármaco usado para la obtención de las imágenes fue el ^{99m}Tc -ECD. Suministrada a los pacientes como una inyección, esta sustancia emite rayos gamma al descomponerse en el interior del organismo. Los rayos gamma son detectados por una cámara Picker Prism 3000 que tomó 180 proyecciones con resolución angular de 2 grados para cada imagen. Las imágenes de secciones del cerebro se reconstruyeron a partir de los datos de proyección usando el algoritmo de retroproyección filtrada en combinación con un filtro de Butterworth para eliminar ruido de alta frecuencia.

Una vez que las imágenes SPECT han sido reconstruidas a partir de la información de las proyecciones obtenidas por la cámara de rayos gamma, éstas deben ser normalizadas espacialmente de manera que podamos aseverar que los vóxeles de diferentes imágenes representan la misma posición anatómica subyacente en el cerebro. Este paso, por tanto, nos permite comparar las intensidades de los vóxeles de distintos sujetos [Salas-Gonzalez *et al.*, 2008b]. La normalización espacial se llevó a cabo mediante el software SPM (en el capítulo 6 se describen ampliamente las etapas de preprocesamiento que realiza este software).

Expertos del hospital “Virgen de las Nieves” etiquetaron posteriormente las imágenes SPECT de forma que quedaron divididas en 4 grupos:

- **NOR (Normales)**. Imágenes de pacientes normales, es decir, que no sufren Alzheimer ni ninguna otra demencia. A estos pacientes comúnmente se les conoce como controles.
 - **AD1 (Posible AD)**. Pacientes con una demencia leve que puede ser considerada una etapa precoz de la enfermedad de Alzheimer.
 - **AD2 (Pobable AD)**. Pacientes que presentan evidencias de sufrir Alzheimer pero aún se encuentran en una etapa temprana de la enfermedad.
 - **AD3 (AD cierto)**. Enfermos claros de Alzheimer que se encuentran en una etapa avanzada de la enfermedad.
-



Figura 3.1: Gamma Cámara Picker Prism 3000

La tabla 3.1 muestra los detalles demográficos de los sujetos usados para esta base de datos.

3.2. Base de datos del centro PET Cartuja

La segunda base de datos usada en esta tesis está formada por 60 imágenes PET de 60 sujetos diferentes proporcionada por el Centro de Diagnostico “PET Cartuja” (CDPC), situado en Sevilla (España). Para más información sobre este centro véase el sitio web www.petcartuja.com.

Las técnicas de diagnóstico PET permiten la diferenciación entre tejido enfermo y sano. Consiste en el registro de imágenes que presentan en 3 dimensiones la distribución orgánica de la molécula más utilizada, la FDG (solución de glucosa marcada con flúor 18). En el cerebro, la distribución de FDG está directamente relacionada con la actividad neuronal y, en el miocardio refleja la actividad cardíaca. Las cámaras PET recogen las señales (radiación) y reproducen en imágenes los procesos detectados.

Tabla 3.1: Detalles demográficos de la base de datos del Hospital Virgen de las Nieves. μ y σ representan la media y la desviación típica respectivamente.

	#	Sex (%)		μ	Age	
		M	F		σ	range
NOR	41	73.17	26.83	71.51	7.99	46-85
AD1	30	36.67	63.33	65.86	13.36	23-81
AD2	22	54.55	45.45	67.22	8.25	46-86
AD3	4	0	100.00	76	9.90	69-83

El protocolo de adquisición seguido en PET-Cartuja se detalla a continuación: La adquisición de las imágenes no comenzó hasta pasados 30 minutos desde la administración del radiofármaco (^{18}F FDG). En este tiempo el paciente permanece en reposo en una habitación en silencio y con iluminación tenue para que el radiofármaco se distribuya adecuadamente por todo el organismo. Se recomienda fijar un tiempo estándar, por ejemplo, 30 minutos, para que los estudios de distintos pacientes a los de distintos controles sean comparables. El paciente se posiciona en decúbito supino. Debe utilizarse un sistema dedicado para apoyar confortablemente la cabeza, fijándola mediante cintas para evitar movimientos involuntarios. El paciente se sitúa acostado en la camilla de la cámara y se desplazará progresivamente por el centro del anillo de la Cámara PET durante un tiempo aproximado de 30 minutos en un estudio de PET cerebral.

Durante este tiempo la cámara PET recoge las señales emitidas por el radiofármaco en todo el cuerpo. Después, un ordenador recoge las señales emitidas y las convierte en imágenes funcionales tridimensionales en los tres planos del espacio (axial, coronal y sagital) y una imagen volumétrica del organismo en tres dimensiones. Los positrones emitidos por el radiofármaco colisionan con los electrones (con carga negativa) de los átomos que componen las moléculas tisulares. La interacción positrón-electrón origina la aparición de un par de fotones con el aniquilamiento de las masas del positrón y electrón. Estos dos fotones presentan una energía de 512KeV cada uno, y se desplazan en la misma dirección y en sentidos opuestos, excitando de forma simultánea 2 detectores de la cámara PET que se encuentran en coincidencia en un ángulo de 180° . Esta detección permite “por coincidencia” la reconstrucción tomográfica tridimensional del organismo que representa la distribución tisular del radiofármaco.

En este caso las imágenes se encuentran etiquetadas formando sólo dos grupos: sujetos sanos (NC) y pacientes con Alzheimer (AD). La base de datos

contiene 47 AD y 13 NC, lo que supone una prevalencia de la clase AD que será tenida en cuenta en los análisis posteriores.

3.3. Base de datos de la iniciativa ADNI

El proyecto ADNI (Alzheimer's Disease Neuroimaging Initiative) fue iniciado en 2003 por el NIA (National Institute on Aging), NIBIB (National Institute of Biomedical Imaging and Bioengineering), el FDA (Food and Drug Administration), compañías farmacéuticas privadas y organizaciones sin ánimo de lucro, como un proyecto conjunto con financiación público-privada y con un presupuesto que asciende a \$ 60 millones de dolares. El primer objetivo de ADNI es probar si se pueden combinar las técnicas de MRI, PET, otros marcadores biológicos y evaluaciones neuropsicológicas y clínicas, para medir la progresión de la enfermedad de Alzheimer en sus primeras etapas (MCI). El descubrimiento de marcadores sensibles y específicos en las etapas más tempranas de la enfermedad se espera que sirva de ayuda a los investigadores y médicos para desarrollar nuevos tratamientos y probar su efectividad, así como para disminuir el tiempo y el coste de las pruebas clínicas.

El investigador principal de la iniciativa ADNI es Michael W. Weiner (Universidad de California, San Francisco). Los pacientes de ADNI fueron seleccionados de más de 50 lugares entre Estados Unidos y Canadá. El objetivo inicial del proyecto ADNI era conseguir que 800 adultos participaran en el proyecto, con edades entre 55 y 90 años, distribuidos en, aproximadamente, 200 sujetos normales de edad avanzada con un seguimiento previsto de 3 años, 400 personas con MCI para un seguimiento de 3 años, y 200 personas con los primeros síntomas de la enfermedad de Alzheimer para un seguimiento de 2 años (véase www.adni-info.org para obtener más detalles sobre los objetivos de este proyecto).

3.3.1. Protocolo de adquisición

Todas las exploraciones FDG PET se tomaron siguiendo un protocolo estandarizado, puesto que un gran número de centros participaron en el estudio y no todos disponían de los mismos recursos materiales para efectuar las exploraciones. Las exploraciones fueron adquiridas según uno de los tres protocolos siguientes:

- **Dinámico.** Tras la la inyección intravenosa de 5.0 ± 0.5 mCi de ^{18}F -FDG se realiza una toma cada 6 minutos durante un periodo de entre 30 y 60

minutos.

- **Estático.** Este protocolo está pensado para los escaners Siemens PET/CT que no tienen la posibilidad de tomar exploraciones dinámicas. Por tanto, se realiza una única toma de 30 minutos de duración comenzando entre 30 y 60 minutos después de administrar el radiofármaco.
- **Cuantitativo.** Consiste en una exploración dinámica del doble de duración que las del protocolo dinámico. Se realizan 33 tomas, repartidas en los 60 minutos siguientes a la administración del radiofármaco. Este protocolo puede ser usado para calcular la tasa absoluta de metabolismo de glucosa, obtenida de la función de entrada del radioisótopo que se mide en las arterias carótidas.

La mayoría de las exploraciones de ADNI fueron tomadas siguiendo el primer protocolo. Los pacientes, a los que se solicitó que ayunaran por lo menos durante las cuatro horas previas a la exploración, permanecieron tumbados relajadamente, con sus ojos abiertos y la estimulación sensorial mínima.

Una vez adquiridas las imágenes, la iniciativa ADNI pretende también minimizar en lo posible, las diferencias en los tipos de imágenes debido a la diferencia entre los escaners usados para obtenerlas. De esta manera, proporciona diferentes conjuntos de datos que se pueden descargar de sus sistemas, ordenados según el tratamiento recibido. El preprocesamiento de las imágenes en ADNI tiene dos objetivos principales. Por un lado trata de hacer más uniforme la información PET disponible y, por otro lado, hace más parecidas las imágenes que proceden de diferentes sistemas.

El preprocesamiento que se sigue en ANDI se estructura en 4 etapas. Si bien, sólo a las imágenes adquiridas con los protocolos dinámico y cuantitativo se le realizan las 4 etapas. Las imágenes adquiridas con protocolo estático sólo necesitan 2 de ellas. A continuación se describe cada una de estas etapas.

1. **Co-registro dinámico:** Una vez los archivos han sido descargados por la Universidad de Michigan, las imágenes Raw son convertidas a un formato de archivo estándar. Las diferentes tomas que pudiera contener el archivo son extraídas y registradas por separado. En la mayoría de los casos, un archivo contiene 6 tomas de 5 minutos tomadas entre 30 y 60 minutos después de la inyección del radiofármaco. Cada toma es co-registrada con respecto a la primera toma (la que se tomó en los 30-35 minutos posteriores a la inyección). Posteriormente las tomas se recombinan formando una nueva imagen dinámica. Estas series de imágenes tienen el mismo número
-

de vóxeles (usualmente, $128 \times 128 \times 63$ vóxeles), el mismo tamaño de vóxel (normalmente $2,0 \times 2,0 \times 2,0$ milímetros) y la misma orientación espacial que la imagen PET original. Los parámetros de la imagen original (número de vóxeles, tamaño, etc) se denomina “espacio ‘nativo’”. En resumen, el co-registro dinámico consigue reducir los efectos de los movimientos del paciente mediante la separación de las secuencias y el co-registrado de las mismas entre sí. Por tanto esta etapa sólo se realiza en caso de que haya varias tomas, es decir, cuando se siguió un protocolo dinámico o uno cuantitativo. Además al final esta etapa, se garantiza que las imágenes se encuentran en el formato estándar DICOM.

2. **Construcción de la imagen promediada:** Esta imagen se genera como un promedio de sólo 6 secuencias de cinco minutos (o las últimas 6 secuencias de los estudios que siguen un protocolo cuantitativo) de la imagen co-registrada que se ha descrito anteriormente. Esto crea una única imagen PET de 30 minutos, todavía en el “espacio nativo”. Como en el caso anterior, sólo las exploraciones de PET adquiridas usando el protocolo dinámico o cuantitativo requieren esta etapa.
 3. **Estandarización de la imagen y del tamaño del vóxel:** La imagen promediada (o la imagen original en caso de estudios de protocolo estático) es redimensionada a un tamaño estándar de $160 \times 160 \times 96$ vóxeles, representando cada vóxel un cubo de 1,5 milímetros de lado. Esta imagen es reorientada de forma que eje anterior-posterior sea paralelo a la línea AC-PC (línea imaginaria que une la comisura anterior y la comisura posterior). Ésto es conocido en LONI como “espacio AC-PC”. Esta imagen estandarizada se usa como imagen de referencia para todos los escáneres de un sujeto (en la base de datos hay muchos escáner de un mismo paciente para realizar un seguimiento: el estudio base, otro a los 6 meses, otro al año, etc). Cada toma individual de cada escáner PET es co-registrada de acuerdo a la imagen de referencia. Haciendo la transformación de la imagen original (datos PET Raw) a un espacio estándar en un paso, sólo se requiere una interpolación de la imagen y así, la degradación de la resolución debida a la interpolación es menor. Posteriormente la intensidad es normalizada usando una mascara específica para cada sujeto de forma que la media de vóxeles dentro de la mascara es exactamente 1. Tanto la reorientación espacial (AC-PC) como la normalización en intensidad son considerados el punto de partida para posteriores análisis. Con una imagen estandarizada, comparar datos PET procedentes de distintos centros con distinta maquinaria es más fácil. Nótese que solamente se realiza una
-

normalización espacial y otra en intensidad, no se hacen deformaciones ni escalados lineales.

4. **Resolución uniforme:** En esta etapa se realiza el suavizado de las imágenes. Cada imagen se filtra con una función de filtro específica (puede ser un filtro no isotrópico) para producir imágenes de una resolución isotrópica uniforme de 8 mm FWHM (la menor resolución usada por un escáner en ADNI). Las imágenes procedente de los escáneres de mayor resolución, obviamente, deben ser suavizadas más que las imágenes de escáneres de baja resolución. Las funciones filtro se determinan a partir de las exploraciones de escáner PET Hoffman phantom que fueron adquiridos durante el proceso de certificación.

En este trabajo ha sido seleccionado un conjunto de 403 imágenes FDG PET de 403 sujetos diferentes (todas las imágenes corresponden a la primera sesión de cada sujeto), adquiridas con escáneres Siemens, General Electric (GE) y Philips PET. Los datos adquiridos con escáneres Siemens HRRT y BioGraph HiRez fueron excluidos debidos a las diferencias en el patrón de toma de imágenes.

3.3.2. Criterios de Etiquetado

Los criterios de elección que se siguieron para aceptar a participantes en el proyecto ADNI se basaron en una serie de entrevistas y test realizados individualmente. Los resultados de los candidatos debían cumplir ciertas condiciones para ser admitidos en el proyecto. A continuación se detallan los criterios de selección de pacientes para cada una de las clases de interés para el estudio:

- Pacientes NC (Normal Control): La puntuación obtenida en el test MMSE debía estar entre 24-30 (ambos inclusive), un CDR de 0, no deprimido, no MCI, y sin demencia. El rango de edad de los pacientes normales debe coincidir con la de los sujetos AD y MCI.
 - Pacientes MCIs: La puntuación obtenida en el test MMSE debía estar entre 24-30 (ambos inclusive), debían presentar quejas por pérdida de memoria, tener una pérdida objetiva de memoria medida en términos de su puntuación en el test de Wechsler Memory Scale Logical Memory II, un CDR de 0.5, ausencia de discapacidades, conducta normal en las actividades de la vida cotidiana y ausencia de demencia.
-

Tabla 3.2: Detalles demográficos de la base de datos ADNI usada en este trabajo. μ y σ representan la media y la desviación típica respectivamente.

	#	Sexo		Edad		Rango
		M	F	μ	σ	
NC	97	60	37	75.97	4.91	62-86
MCI	188	122	66	75.12	7.22	55-89
MCIc	23	18	5	73.97	7.35	57-85
AD	95	57	38	75.72	7.40	55-88

- Pacientes AD: La puntuación obtenida en el test MMSE debía estar entre 24-30 (ambos inclusive), debía presentar un CDR de 0.5 ó 1.0, y satisfacer los criterios de NINCDS/ADRDA que definen un AD probable.

Consecuentemente, los datos FDG PET se separaron en 3 clases diferentes: pacientes normales de control (NC), pacientes con afección cognitiva leve (MCI) y pacientes con la enfermedad de Alzheimer (AD). Además, la iniciativa ADNI proporciona información acerca de aquellos pacientes MCI que permanecen estables durante 2 años y aquellos que evolucionaron hacia AD, permitiéndonos distinguir entre MCI y MCI converters (MCIc). La composición de esta base de datos se muestra en la tabla 3.2.

CAPÍTULO

4

LA ENFERMEDAD DE ALZHEIMER

La enfermedad de Alzheimer es la causa más común de demencia, por ello, el desarrollo de nuevos métodos para el análisis de imágenes tomográficas que permitan mejorar el diagnóstico precoz, tendrá especial importancia en el caso de AD. En este capítulo se ofrece una visión general sobre la enfermedad, sus causas y los test y escalas usados tradicionalmente para el diagnóstico.

4.1. La enfermedad de Alzheimer

La enfermedad de Alzheimer (también llamada mal de Alzheimer o simplemente Alzheimer) recibe su nombre en honor al médico alemán Alois Alzheimer quien, en 1906, diagnosticó la enfermedad por primera vez. La paciente a la que fue diagnosticada la enfermedad fue Auguste Deter, hospitalizada en 1901 cuando mostraba síntomas agudos de la enfermedad como habilidades de memoria muy reducidas, pérdida de orientación, etc. Tras su muerte en 1906, Alzheimer diseccionó su cerebro y encontró placas de amiloides y ovillos neurofibrilares que hoy en día son considerados los signos característicos de AD. Estos descubrimientos fueron publicados en 1907 [Alzheimer, 1907]. La enfermedad de Alzheimer es una enfermedad lenta, comienza con leves problemas de memoria y termina con daños cerebrales severos. La evolución de la enfermedad y la rapidez con la que avanza depende del paciente, si bien, los pacientes de Alzheimer viven entre 8 y 10 de media después de que la enfermedad les fuera diagnosticada. La Organización Mundial de la Salud estimó que en 2005 el 0,379 % de las personas a nivel mundial tenían demencia y que la prevalencia aumentaría a un 0,441 % en 2015 y a un 0,556 % en 2030 [Organization, 2006].



Figura 4.1: Alois Alzheimer

La principal característica de la enfermedad de Alzheimer en su etapa inicial son los olvidos. Los pacientes no sólo olvidarán donde dejaron las cosas sino que las colocarán en sitios inusuales. La pérdida de la habilidad de calcular el tiempo transcurrido es otro indicador de la enfermedad. Perder el sentido

Tabla 4.1: Número de afectados de Alzheimer por cada mil habitantes en función de su edad [Bermejo-Pareja *et al.*, 2008]

Edad	Afectados por cada 1000 habitantes
65 - 69	3
70 - 74	6
75 - 79	9
80 - 84	23
85 - 89	40
90 o más	69

del tiempo y de la orientación hace que los pacientes tengan dificultades para desplazarse y que lleguen incluso a perderse, especialmente en sitios con los que están menos familiarizados. Las tareas cotidianas raramente se ven afectadas, en cambio tareas abstractas como hacer cálculos, etc se convierten en actividades complejas que los pacientes se vuelven incapaces de resolver. La degradación de estas funciones es tan gradual que puede tardar muchos años en ser percibida por los familiares [Stoeckel, 2003]. Con el avance de la enfermedad el paciente comienza incluso a perderse en ambientes familiares y se muestra incapaz de aprender cualquier cosa nueva. Los pacientes se vuelven más inquietos y agitados y su comportamiento empieza a ser un problema. Algunos pacientes llegan a ser agresivos, otros en cambio se vuelven apáticos y casi todos sufren alucinaciones y paranoias. En la última etapa de la enfermedad los pacientes son incapaces de reconocer cualquier objeto familiar y pierden totalmente el sentido del tiempo y de la orientación. Son incapaces de moverse por sí solos y necesitan ayuda para realizar prácticamente cualquier tarea. La capacidad de comunicarse también se ve seriamente mermada llegando incluso a no poder articular palabras.

El paciente con Alzheimer no muere por la enfermedad, sino por infecciones secundarias como una llaga de presión o úlcera de decúbito, lesiones que se producen cuando una persona permanece mucho tiempo en la misma posición.

4.1.1. Causas

Actualmente no existe un consenso en la comunidad científica sobre las causas de la enfermedad de Alzheimer. Ni siquiera hay consenso sobre si hay o no causas y por ello se habla de factores de riesgo.

En algunos casos, menos del 2% la enfermedad es causada por mutaciones genéticas en la familia. En estos casos los síntomas comienzan antes de los 60 años y avanzan rápidamente. Todas las mutaciones conocidas conllevan la sobreproducción de una proteína que destruye las células nerviosas. Además se sabe que algunos factores genéticos aumentan las posibilidades de que una persona desarrolle la enfermedad. Concretamente, los portadores de una variante del gen *ApoE*, que fomenta la creación de proteínas perjudiciales tienen más probabilidades de desarrollar AD [Langbaum *et al.*, 2009; Filippini *et al.*, 2009].

La edad es el factor de riesgo más importante. La cantidad de gente mayor de 65 años que padece la enfermedad se dobla cada 5 años. Por otro lado, las mujeres y las personas con menor nivel de estudios tienen más probabilidades de sufrir esta demencia. Otros factores como el entorno o la dieta parecen no tener influencia, no obstante, todavía se siguen realizando estudios al respecto.

4.1.2. Anatomía patológica

La única forma de obtener un diagnóstico exacto de AD es mediante un examen post-muerte donde se busque la presencia de las alteraciones histológicas que produce esta demencia. Estas alteraciones pueden ser de tres tipos: ovillos neurofibrilares, vacuolas intracito-plasmáticas y placas seniles, conocidas también como placas de amiloide [Pérez, 2010].

Los ovillos neurofibrilares son estructuras intraneuronales cuya presencia se comprueba mediante microscopia fotónica, tras impregnación argéntica de los citoplasmas de las neuronas. Tienen la forma de un haz de fibrillas anómalas que pueden tomar el aspecto de llamas o de bolas que, a menudo, deforman la neurona y vuelven el núcleo excéntrico. En microscopia electrónica, esta lesión corresponde a un enmarañamiento de filamentos de 10 nm de espesor, organizados por pares, según una configuración helicoidal con un paso de 80 nm. En las neuronas piramidales del hipocampo y los ovillos neurofibrilares es posible encontrar vacuolas intracito-plasmáticas de 4 ó 5 μm de diámetro, que contienen un gránulo central eosinófilo.

Las placas seniles son estructuras redondeadas situadas exclusivamente en las zonas corticales y con predominio en la segunda y la tercera capa. Se componen de un centro amiloide, que se pone perfectamente de manifiesto mediante el rojo congo, rodeado de una corona de prolongaciones nerviosas en degeneración. Mediante un microscopio electrónico es posible observar como el centro de las placas (sustancia beta-amiloide) está constituido por un enmarañamiento de filamentos rectilíneos con una configuración beta plegada de 7 a 10 nm de diámetro. Por regla general, las placas seniles se asocian a una importante

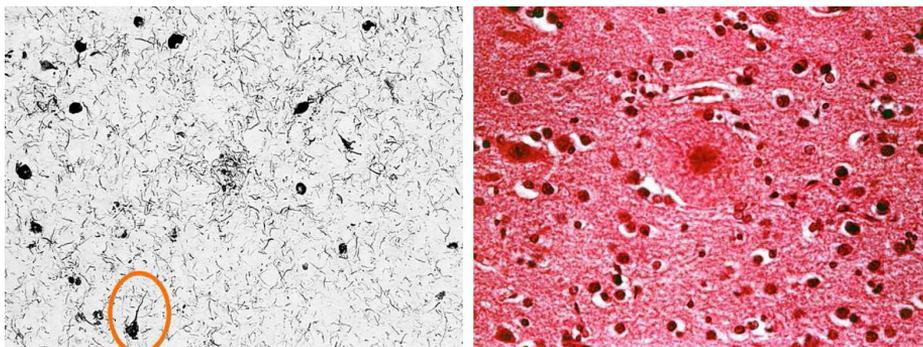


Figura 4.2: Las lesiones específicas de la enfermedad de Alzheimer sólo pueden observarse con microscopio. Los ovillos neurofibrilares (izquierda) que se caracteriza por la agregación anormal de proteínas muy específicas (proteínas Tau) en las neuronas. Las placas seniles (en el centro de la imagen de la derecha) están constituidas por un depósito del péptido beta-amiloide, del que hasta la fecha no se conoce su función exacta

densidad de ovillos neurofibrilares. Estudios recientes han demostrado que estas placas no están constituidas por materia inerte sino que presentan actividades metabólicas múltiples para diversos sistemas de neuromediadores [Pérez, 2010].

En la figura 4.2 se muestra el aspecto de las placas seniles y de ovillos neurofibrilares. Estas lesiones elementales predominan en las zonas corticales y en las áreas asociativas del cerebro, en particular las regiones posteriores (encrucijada parietotemporoccipital y la parte interior y externa de los lóbulos temporales). Se acompañan de una despoblación neuronal con atrofia cortical. La alteración de ciertos núcleos grises centrales, como el núcleo basal de Meynert, indica sin embargo, una participación subcortical abiotrófica del proceso, cuya importancia sigue en la actualidad en discusión. En definitiva, las características anatómicas de la enfermedad no se basan tanto en la naturaleza misma de las lesiones, cuya especificidad es insignificante (las placas seniles y de ovillos neurofibrilares se encuentran, asimismo, en la trisomía 21, en el Parkinson postencefalítico, en la demencia de los boxeadores y, sobre todo, en individuos de edad avanzada sin demencia), sino especialmente en el número y en la difusión a todas las capas corticales de estas alteraciones elementales. Además, aunque las anomalías histológicas son cualitativamente idénticas en las formas preseniles y seniles de la enfermedad, la topografía de la atrofia y la intensidad de las alteraciones

diferirían notablemente en estas dos formas.

4.2. Técnicas de diagnóstico de AD

Como se ha visto en la sección 1.3.2 un diagnóstico temprano de la enfermedad es fundamental para mejorar y alargar la vida del paciente. Existen tres niveles en el camino hacia el diagnóstico precoz y seguro de AD:

- Diagnóstico basado en tests y entrevistas personales. Es la forma de diagnóstico tradicional y durante mucho tiempo, fue la única. Los cuestionarios o escalas fueron diseñados para cuantificar determinadas funciones cognitivas, es decir, no establecen un diagnóstico, sino que cuantifican la severidad de la alteración de determinadas áreas intelectuales, siendo particularmente valiosos para discriminar entre envejecimiento normal y demencias leves.
- Examen visual de imágenes funcionales. Tras la llegada de las técnicas de medicina nuclear y su gran potencial para obtener imágenes funcionales del cerebro, los médicos suelen ayudarse de estas imágenes para diagnosticar AD. Normalmente la presencia de hipoperfusión en determinadas zonas del cerebro previamente establecidas es indicativo de AD.
- Sistemas informáticos de ayuda al diagnóstico. El último escalón lo componen los sistemas CAD. Se trata de sistemas informáticos que hacen uso de técnicas de aprendizaje estadístico y de bases de datos de imágenes etiquetadas para diagnosticar AD de forma automática, dada una imagen cerebral del paciente. Es importante resaltar que estos sistemas suponen *sólo* una ayuda al diagnóstico que siempre deberá ser emitido por un médico en base a la historia clínica del paciente y de acuerdo a los criterios establecidos al respecto.

El diagnóstico de la demencia se basa fundamentalmente en la evaluación clínica y ésta requiere una exhaustiva evaluación de la función cognitiva, en concreto memoria, atención, percepción, lenguaje, praxias y gnosias. La evaluación neuropsicológica puede subdividirse en dos niveles de complejidad: *i*) un primer nivel que consiste en el uso de test breves, estandarizados y sencillos como el examen del estado mental mínimo (MMSE), que permitan alcanzar el diagnóstico de demencia; y *ii*) un segundo nivel de mayor complejidad en el que se refina la evaluación de la severidad del deterioro, al tiempo que se establecen

los dominios de la función cognitiva que se hallan afectados. Existen diferentes escalas que otorgan un valor estandarizado en función del grado de afectación funcional, como la escala de deterioro global (GDS) o la clasificación clínica de las demencias (CDR). En general se intenta que estas escalas permitan clasificar la demencia según los criterios clínicos clásicos: demencia leve, moderada o severa.

4.2.1. Examen del estado mental mínimo

El examen del estado mental mínimo (MMSE, del inglés Mini-Mental State Examination) es uno de los test más utilizados para la evaluación de diferentes demencias. Es un test que tiene alta dependencia del lenguaje y consta de varios ítems relacionados con la atención. Cada ítem tiene una puntuación, llegando a un total de 30 puntos. En la práctica diaria una puntuación menor de 24 sugiere demencia, entre 23-21 una demencia leve, entre 20-11 una demencia moderada y menor de 10 de una demencia severa. Para poder efectuar el MMSE es necesario que el paciente se encuentre despierto y lúcido. En la demencia por enfermedad de Alzheimer la tasa promedio anual de cambio en la puntuación del MMSE es de 2-5 puntos por año, por lo que el test muestra su utilidad para el seguimiento de pacientes dementes. El MMSE tiene baja sensibilidad para el diagnóstico de deterioro cognitivo leve, la demencia frontal-subcortical y el déficit focal cognitivo. Las características esenciales que se evalúan son:

- Capacidad de atención, concentración y memoria.
- Capacidad de abstracción (cálculo).
- Capacidad de lenguaje y percepción viso-espacial.
- Orientación espacio-tiempo.
- Capacidad para seguir instrucciones básicas.

Esta prueba proporciona un instrumento para detectar el de deterioro cognitivo y se puede realizar en poco tiempo. Según sus autores, esto es especialmente importante en el diagnóstico de la demencia, ya que el paciente demente se cansa rápidamente y deja de mostrarse colaborador.

4.2.2. Escala de deterioro global

La escala de deterioro global (GDS, del inglés Global Deterioration Scale) [Sheikh & Yesavage, 1986] establece siete estadios posibles (véase tabla 4.2).

Tabla 4.2: Diferentes estadios que establece la escala de deterioro global

Estadio	Déficit cognitivo	Características clínicas
1	Inexistente	Normal
2	Muy leve	Olvidos subjetivos pero con exploración normal
3	Leve	Dificultad en el trabajo, en la comunicación verbal o al recorrer lugares poco familiares; detectable por la familia; déficit sutil de memoria en la exploración
4	Moderado	Disminución de la capacidad para viajar, para contar o para recordar acontecimientos recientes
5	Moderadamente grave	Necesita ayuda para elegir la ropa; desorientación en el tiempo o espacio; recuerda peor el nombre de sus nietos
6	Grave	Necesita supervisión para comer y asearse, posible incontinencia; desorientación en el tiempo, espacio y posiblemente en identidad
7	Muy grave	Perdida importante de la capacidad verbal, incontinencia y rigidez motora

Esta escala define cada estadio en términos operacionales y en base a un deterioro supuestamente homogéneo. Sin embargo, dado que la secuencia de aparición de los síntomas es a menudo variable, se ha argumentado que la inclusión de un paciente en un estadio de acuerdo a un criterio rígido podría conducir a errores. No obstante se trata de una de las escalas más completas, simples y útiles para la estimación de la severidad de la demencia. El CAED sugirió en 1997 la utilización de esta escala para la gradación del síndrome demencial de la enfermedad de Alzheimer acompañado del Instrumento de Evaluación Funcional para Enfermedad de Alzheimer (FAST, del inglés Functional Assessment tool for Alzheimer's disease).

4.2.3. Clasificación clínica de la demencia

La evaluación de las demencias que no son enfermedad de Alzheimer se realiza a través de la clasificación clínica de la demencia (CDR, del inglés, Clinical Dementia Rating) [Morris, 1993] que es más general. Su escala establece cinco estadios posibles:

- 0 → Normal
- 0,5 → Demencia cuestionable
- 1 → Demencia leve
- 2 → Demencia moderada
- 3 → Demencia severa

La estimación se realiza en base al rendimiento del sujeto en seis modalidades de tipo cognitivo y funcional. Estas modalidades son: memoria, orientación, razonamiento, actividades sociolaborales, actividades recreativas (hobbies o pasatiempos), y cuidado personal.

4.2.4. Escala de evaluación para la enfermedad de Alzheimer

En 1984 con la aparición de la escala de evaluación para la enfermedad de Alzheimer (ADAS, del inglés Alzheimer's Disease Assessment Scale) fue posible contar con un instrumento fiable diseñado especialmente para la enfermedad de Alzheimer y capaz de medir puntualmente los síntomas característicos de la misma así como su progresión a estadios más avanzados.

El ADAS es un test que evalúa rendimiento y que consta de 21 ítems segregados en dos subescalas: cognitiva (ADAS-Cog) y conductual (ADAS-Noncog), de 11 y 10 ítems respectivamente. En la práctica se ha hecho muy popular el uso del ADAS-Cog, mientras que el ADAS-Noncog se ha reemplazado por escalas conductuales más generales como Inventario Neuropsiquiátrico o NPI-Q (Neuro-Psiquiatric inventory questionnaire). Este último es un instrumento de aplicación relativamente breve que mide una serie de síntomas habituales en función de la frecuencia de aparición y la intensidad con que aparecen, aportando también una puntuación global que es la suma de las anteriores.

Parte II

Fundamentos teóricos

CAPÍTULO

5

MÉTODOS DE CLASIFICACIÓN

Los sistemas de ayuda al diagnóstico por ordenador basados en enfoques multivariados se basan en el uso de *métodos de clasificación*. Un método de clasificación es un algoritmo que agrupa (o discrimina) objetos, descritos mediante un vector de atributos, asignándolos a clases previamente definidas. Un clasificador es un algoritmo de aprendizaje automático supervisado que emplea un conjunto de objetos etiquetados (ejemplos) para construir una función de clasificación que asigne una etiqueta a un objeto del que *a priori* se desconoce su etiqueta. Tras una introducción sobre aprendizaje automático y métodos de clasificación, en este capítulo veremos una breve descripción de los clasificadores usados en este trabajo y los principales métodos para evaluar el rendimiento de los mismos.

5.1. Introducción

El aprendizaje automático se define como la capacidad de ciertos sistemas para extraer la información subyacente en un conjunto de datos que son vistos como ejemplos del comportamiento de ciertas variables observables. El aprendizaje supervisado es una rama del aprendizaje automático que calcula una función para establecer una correspondencia entre las entradas y las salidas deseadas del sistema. Un ejemplo de este tipo de algoritmos de aprendizaje supervisado son los métodos de clasificación en los que el sistema de aprendizaje trata de etiquetar (clasificar) una serie de objetos. En este caso, los datos de ejemplo son objetos previamente etiquetados, es decir, objetos cuya clase es conocida. Los sistemas de diagnóstico por ordenador usan un tipo especial de aprendizaje supervisado en el que las etiquetas son binarias, esto es, sólo hay dos etiquetas posibles: paciente sano y paciente enfermo. El proceso de aprendizaje requiere por tanto de un conjunto de objetos previamente etiquetados y cuyo error de etiquetado se asumirá despreciable.

En el caso de sistemas de diagnóstico para enfermedades neurológicas, los objetos se representan con vectores de características, cada uno de cuales es construido a partir de la imagen funcional del cerebro de un paciente. La construcción de estos vectores de características a partir de las imágenes SPECT o PET es una cuestión que influye de manera determinante en el funcionamiento general del algoritmo y será abordada con detalle en la sección 5.4 y en los capítulos 7 y 8. En muchas ocasiones, el conjunto de vectores de características queda representado de una manera más simple en un espacio diferente al original. En estos casos, la eficiencia del clasificador mejora cuando se realiza la transformación oportuna, a través de las funciones conocidas como *kernels*. El criterio a seguir lo marcará el *Teorema de Mercer* que se verá en la sección 5.2.

El conjunto de datos experimentales o muestras en el proceso de clasificación estará formado por un conjunto de *vectores de características* $\mathbf{x}_i \in \mathbb{R}^m$, $i = 1, \dots, n$, siendo m la dimensión del *espacio de características* \mathcal{H} . Este conjunto de vectores de características se dividirá en 2 subconjuntos: datos de entrenamiento y datos de test.

- Diremos que un vector de características \mathbf{x}_i es un *vector de entrenamiento* si pertenece al subconjunto de entrenamiento $\mathcal{X} \subset \mathbb{R}^m$. Expresado matemáticamente, $\mathbf{x}_i \in \mathcal{X}$.
 - Diremos que un vector de características $\bar{\mathbf{x}}_j$ es un *vector de test* si pertenece al subconjunto de test $\mathcal{Y} \subset \mathbb{R}^m$. Expresado matemáticamente, $\bar{\mathbf{x}}_j \in \mathcal{Y}$.
-

- La unión de estos dos subconjuntos formará el espacio de características \mathcal{H} , es decir, $\mathcal{H} = \mathcal{X} \cup \mathcal{Y}$. En nuestro caso concreto, asumiremos que \mathcal{H} es sencillamente \mathbb{R}^m .

La elección de los conjuntos de entrenamiento y test es un problema ampliamente estudiado en la literatura. Una condición necesaria es que los conjuntos sean independientes, es decir, que no existan elementos del conjunto de test que hayan sido utilizados en el entrenamiento del clasificador:

$$\mathcal{X} \cap \mathcal{Y} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \cap \{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_n\} = \emptyset \quad (5.1)$$

Se define un clasificador como una función $f = f(\mathbf{x}_i, \omega)$, dependiente de unos parámetros ω y construida a partir de los datos de entrenamiento n -dimensionales $\mathbf{x}_i \in \mathcal{X}$ y sus correspondientes etiquetas y_i . Esta función clasifica un nuevo vector de test $\bar{\mathbf{x}}_j \in \mathcal{Y}$, asignándole un valor $z_j \in \{\pm 1\}$ correspondiente a cada clase:

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \{\pm 1\} \\ \mathbf{x}_j \in \mathcal{Y} &\longmapsto f(\mathbf{x}_j, \omega) = z_j \end{aligned} \quad (5.2)$$

El proceso de clasificación, por tanto, se divide en dos etapas: entrenamiento y test. En la etapa de entrenamiento se define el clasificador de acuerdo con un conjunto de vectores de características de entrenamiento $\mathbf{x}_i \in \mathcal{X}$, cuyas etiquetas son conocidas. La forma exacta del clasificador dependerá del modelo propuesto para la tarea de clasificación, y los parámetros desconocidos del modelo serán estimados empleando los patrones de entrenamiento $\mathbf{x}_i \in \mathcal{X}$, $i = 1, \dots, l$. El algoritmo de aprendizaje automático consistirá en estimar los parámetros ω , de manera que el error de clasificación sobre el conjunto de entrenamiento sea mínimo. Una vez definido, el clasificador se emplea para establecer categorías sobre muestras desconocidas pertenecientes al conjunto de test.

Uno de los aspectos interesantes a evaluar sobre la clasificación es cómo los resultados pueden ser generalizados a un conjunto de datos independientes. La *generalización* es un concepto central en el diseño de clasificadores y hace referencia a la capacidad de un clasificador construido a partir de un conjunto muestral de entrenamiento determinado, para describir la estructura subyacente de la población total, y no de la muestra concreta. Así, un clasificador con buena capacidad de generalización será capaz de operar correctamente con nuevos datos independientes.

5.2. Métodos Kernel

La tarea de clasificación requiere introducir el concepto de similitud. De alguna manera, al clasificar se comparan los datos \mathbf{x} cuyas etiquetas son conocidas, con nuevos patrones \mathbf{x}' , y se categoriza el nuevo patrón en función de su similitud con los datos conocidos. Matemáticamente, podemos considerar que una medida de la similitud viene dada por una función:

$$\begin{aligned} k : \mathcal{H} \times \mathcal{H} &\longrightarrow \mathbb{R} \\ (\mathbf{x}, \mathbf{x}') &\longmapsto k(\mathbf{x}, \mathbf{x}') \end{aligned} \quad (5.3)$$

es decir, una función que, dadas dos muestras del espacio de características \mathbf{x} y \mathbf{x}' , devuelve un número real que caracteriza su similitud. Es natural asumir que no existe diferencia al comparar la similitud entre \mathbf{x} y \mathbf{x}' que entre \mathbf{x}' y \mathbf{x} , por lo que asumiremos que esta función es simétrica, es decir, $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$, para todo $\mathbf{x}, \mathbf{x}' \in \mathcal{H}$. Un ejemplo sencillo de una función de este tipo es el producto escalar o producto interno (dado que lleva asociada una noción de distancia, puede verse como una medida de similitud) que se define como:

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}' = \sum_{i=1}^m x_i x'_i \quad (5.4)$$

con $\mathbf{x} = (x_1, \dots, x_m)$, $\mathbf{x}' = (x'_1, \dots, x'_m)$ y $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^m$.

Estas ideas se pueden extrapolar a espacios más generales que \mathbb{R}^m , en los que también es posible definir un producto interno. Estos espacios, denominados espacios de Hilbert, son análogos al espacio Euclídeo en tanto que se puede definir en ellos los conceptos de longitud, distancia y ángulo, pero poseen características más generales. Son la generalización del espacio Euclídeo, reproduciendo el álgebra vectorial de manera abstracta e incluyendo una noción de completitud.

Es posible que la noción de similitud se defina de forma más apropiada en otro espacio V diferente al espacio de características original \mathcal{H} . Supondremos que este espacio V es un espacio de Hilbert, con su correspondiente producto interno. En primer lugar es necesario construir una función de mapeo, φ , que lleve los vectores de características del espacio original, \mathcal{H} , al nuevo espacio V :

$$\begin{aligned} \varphi : \mathcal{H} &\longrightarrow V \\ \mathbf{x} &\longmapsto \mathbf{v} = \varphi(\mathbf{x}) \end{aligned} \quad (5.5)$$

Teniendo en cuenta este mapeo, el producto escalar se expresa como:

$$k(\mathbf{v}, \mathbf{v}') = \mathbf{v} \cdot \mathbf{v}' = \varphi(\mathbf{x}) \cdot \varphi(\mathbf{x}') = k(\varphi(\mathbf{x}), \varphi(\mathbf{x}')) \quad (5.6)$$

Por tanto, podemos definir un *kernel* como una función que puede ser representada por un producto escalar en algún espacio de Hilbert. Su uso permitirá seguir trabajando con las nociones habituales de distancia, longitud y ángulo, y las herramientas sencillas empleadas para trabajar con ellas (álgebra lineal, geometría analítica,...) pero abriendo la posibilidad de utilizar un amplio rango de diferentes medidas de la similitud, gracias a la libertad a la hora de elegir el mapeo φ . Las condiciones generales para la existencia de tal mapeo vienen dadas por el Teorema de Mercer.

Teorema de Mercer. Sea $\mathbf{x} \in \mathbb{R}^m$ y φ una función de mapeo

$$\begin{aligned} \varphi : \mathbb{R}^m &\longrightarrow V \\ \mathbf{x} &\longmapsto \mathbf{v} = \varphi(\mathbf{x}) \end{aligned} \quad (5.7)$$

donde V es un espacio de Hilbert. Entonces, la operación del producto interno tiene una representación equivalente

$$\sum_i \varphi_i(\mathbf{x})\varphi_i(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}')$$

donde $\varphi_i(\mathbf{x})$ es la componente i del mapeo $\varphi(\mathbf{x})$ de \mathbf{x} y $k(\mathbf{x}, \mathbf{x}')$ es una función simétrica que satisface la siguiente condición:

$$\int k(\mathbf{x}, \mathbf{x}')g(\mathbf{x})g(\mathbf{x}')d\mu(\mathbf{x})d\mu(\mathbf{x}') \geq 0 \quad (5.8)$$

para cualquier función de cuadrado integrable $g \in L_2$, es decir, tal que

$$\int g(\mathbf{x})^2 d\mu(\mathbf{x}') < +\infty \quad (5.9)$$

donde $d\mu(\mathbf{x}')$ es la medida de Borel-Lebesgue.

La afirmación contraria del teorema es también cierta, es decir, para cualquier $k(\mathbf{x}, \mathbf{x}')$ que satisfaga (5.8) y (5.9) existe un espacio en el cual k defina un producto interno. Puesto que nosotros partimos de una representación del conjunto de datos en un espacio vectorial con producto interno, se puede entender el mapeo (5.7) como un isomorfismo entre espacios de Hilbert, por lo tanto es siempre posible definir un mapeo lineal uno-a-uno entre espacios que preserve la estructura de producto interno. En estos casos, la definición de la función kernel

se conoce como *kernel trick*, en la que en lugar de construir una función kernel, se establece una transformación entre funciones kernel. Más aun, estaremos interesados por funciones g diferenciables, por lo que la medida de Borel-Lebesgue de (5.8) y (5.9) se puede entender como la medida normal de la integral de Riemann. Sin embargo, existen construcciones mucho más generales que permiten usar estas técnicas para casos en los que el espacio de características original no sea un espacio vectorial [Vapnik, 1998].

Gracias al teorema de Mercer podemos estar seguros de se puede establecer un mapeo a un nuevo espacio que permita definir una noción de similitud. Normalmente este nuevo espacio implicará un aumento de la dimensión del espacio de características, para proporcionar un marco en el que las clases se separen más fácilmente. Lo que el teorema de Mercer no revela sin embargo es cómo encontrar este espacio. Es decir, no tenemos una herramienta general para construir el mapeo $\varphi(\cdot)$ una vez que conocemos el producto interno del correspondiente espacio.

En la práctica, ha resultado que los mapeos no lineales producen estructuras suficientemente interesantes para resolver problemas complejos, siendo algunos ejemplos típicos de kernels usados en aplicaciones de reconocimiento de patrones:

- Polinómicos:

$$k(\mathbf{x}, \mathbf{x}') = [\gamma(\mathbf{x} \cdot \mathbf{x}') + c]^d. \quad (5.10)$$

- Funciones de base radial (RBF):

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2). \quad (5.11)$$

- Tangente hiperbólica:

$$k(\mathbf{x}, \mathbf{x}') = \tanh(\gamma(\mathbf{x} \cdot \mathbf{x}') + c). \quad (5.12)$$

para valores apropiados de γ y c de modo que las condiciones de Mercer se cumplan.

5.3. Clasificadores estadísticos

Como hemos visto en las secciones anteriores, un clasificador binario calcula una función, f , conocida como función de decisión que asigna una etiqueta binaria, $C_i, i \in 1, 2$ a cada vector de características, \mathbf{x}_i (véase ecuación 5.2). Un

clasificador lineal es aquel cuya función de decisión se basa en una combinación lineal de los vectores de entrada:

$$f_{lineal}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w} + \mathbf{w}_0 \quad (5.13)$$

donde el vector \mathbf{w} y el escalar \mathbf{w}_0 forman un conjunto de $m + 1$ parámetros que han de ser estimados. Existe un amplio rango de algoritmos de descenso de gradiente para modificar los parámetros de un clasificador de manera que se reduzca en cierta medida del error, dando lugar un nuevo campo, denominado reconocimiento estadístico de patrones, que se encarga de su estudio. Para nuestros propósitos empleamos un conjunto de clasificadores clásicos que se definen en base a una minimización de la función de coste.

Una elección apropiada de un clasificador debe tener en cuenta la distribución de las clases en el espacio de características. Por ejemplo, un clasificador lineal que defina un hiperplano en el espacio de características no será apropiado para separar dos clases representadas por dos círculos concéntricos. Sin embargo, la elección de un clasificador no lineal complejo no siempre garantiza obtener los mejores resultados. De hecho, cuanto mayor sea la complejidad de la función que define el clasificador, más parámetros habrán de ser estimados para definirlo, lo que puede ser problemático sobre todo si se cuenta con un número reducido de vectores de entrenamiento.

En esta sección se describirán brevemente los 7 clasificadores usados en este trabajo, haciendo especial hincapié en las máquinas de vectores soporte por ser el clasificador que mejores resultados proporciona.

5.3.1. Clasificador de la media más cercana

El clasificador de la Media más Cercana (NM, del inglés Nearest Mean) usa los vectores de características medios μ_1 y μ_2 de las muestras de ambas clases C_1 y C_2 , definidos como:

$$\mu_j = \frac{1}{N_j} \sum_{\{i|y_i \in C_j\}} \mathbf{x}_i, \quad j = 1, 2 \quad (5.14)$$

donde N_j es el número de vectores de características \mathbf{x}_i cuya etiqueta y_j pertenece a la clase C_j . En el caso de diagnóstico de la enfermedades neurológicas, las dos clases posibles son *sano* y *enfermo*.

El modelo del clasificador de media más cercana asigna a un vector de características \mathbf{x}' de una muestra desconocida la clase o etiqueta con vector media

más cercano. El clasificador por lo tanto se define como:

$$f_{\text{nm}}(\mathbf{x}') = (\mathbf{x}' - \mu_2)^T(\mathbf{x}' - \mu_2) - (\mathbf{x}' - \mu_1)^T(\mathbf{x}' - \mu_1), \quad (5.15)$$

donde la muestra se etiqueta como *sano* si $f_{\text{nm}}(\mathbf{x}') > 0$ o como *enfermo* si $f_{\text{nm}}(\mathbf{x}') < 0$. Vemos como en (5.15) interviene el producto interno a través de producto de matrices $(\mathbf{x}' - \mu_i)^T(\mathbf{x}' - \mu_i)$, lo que supone un ejemplo de criterio de similitud definido a través de un concepto de distancia.

El clasificador de media más cercana es el clasificador óptimo en el caso de que las clases tengan la misma distribución estrictamente decreciente y simétrica alrededor de la media, y la misma varianza en todas las direcciones. Este es el caso de la distribución gaussiana o normal. Además, las medias de las poblaciones de las clases son conocidas, y no tienen que ser estimadas a través de las medias muestrales (5.15).

5.3.2. Clasificador lineal de Fisher

El clasificador lineal de Fisher (FLD, del inglés Fisher Linear Discriminant) [Fukunaga, 1990] es una extensión del clasificador de media más cercana que tiene en cuenta la forma de la distribución de los vectores de características.

El clasificador de Fisher se define como:

$$f_{\text{FL}}(\mathbf{x}') = (\mathbf{x}' - \mu_2)^T \mathbf{C}^{-1}(\mathbf{x}' - \mu_2) - (\mathbf{x}' - \mu_1)^T \mathbf{C}^{-1}(\mathbf{x}' - \mu_1) \quad (5.16)$$

donde \mathbf{C} es la matriz de covarianza de la distribución de las muestras que se asume idéntica en ambas clases por razones computacionales.

Si el número de muestras de entrenamiento es menor que la dimensión del espacio de características, la matriz de covarianza \mathbf{C} es a una matriz singular, por lo que no puede ser invertida. En ese caso, se pueden seguir diferentes estrategias, como asumir restricciones en la matriz de covarianza (una matriz de covarianza proporcional a la identidad o diagonal) o reemplazar \mathbf{C}^{-1} por su pseudo-inversa, por ejemplo usando una descomposición de valor singular para obtenerla (véase [Fukunaga, 1990]).

5.3.3. Clasificador Bayesiano Ingenuo

El clasificador Bayesiano Ingenuo (NB, del inglés Naive Bayes) está basado en la evaluación la función de probabilidad *a posteriori* [Fukunaga, 1990]. A

continuación se describe el caso general en el que puede haber más de 2 clases. Sean $C_l, l = 1, \dots, L$ las etiquetas para las L clases entre las que distingue el clasificador y \mathbf{x}_i el vector de características correspondiente a una imagen funcional. La función de probabilidad *a posteriori* de C_l dado \mathbf{x}_i es definida como

$$P(C_l|\mathbf{x}_i) = \frac{p(\mathbf{x}_i|C_l)P(C_l)}{P(\mathbf{x}_i)} \quad (5.17)$$

donde $P(C_l)$ es la probabilidad *a priori*, $p(\mathbf{x}_i|C_l)$ es la función de densidad de probabilidad condicional de \mathbf{x}_i dado C_l y $p(\mathbf{x}_i)$ es la densidad de la mezcla. En la practica, el denominador es constante y no depende de C_l por lo que sólo se tiene en cuenta el numerador. El máximo de la regla de decisión *a posteriori* para el clasificador de Bayes es definido como

$$p(\mathbf{x}_i|C_l)P(C_l) = \max_j \{p(\mathbf{x}_i|C_j)P(C_j)\}, \mathbf{x}_i \in C_l \quad (5.18)$$

El vector de características \mathbf{x}_i es clasificado como perteneciente a la clase C_l cuya probabilidad *a posteriori* dado \mathbf{x}_i es mayor. Las densidades dentro de cada clase son normalmente modeladas como distribuciones normales

$$P(C_l|\mathbf{x}_i) = \frac{1}{(2\pi)^{\frac{m}{2}} |\boldsymbol{\Sigma}_l|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(\mathbf{x}_i - \mathbf{M}_l)^t \boldsymbol{\Sigma}_l^{-1} (\mathbf{x}_i - \mathbf{M}_l)\right\} \quad (5.19)$$

donde \mathbf{M}_l y $\boldsymbol{\Sigma}_l$ son la media y la matriz de covarianza de la clase C_l respectivamente y sus expresiones son

$$\mathbf{M}_l = \frac{1}{N_l} \sum_{j=1}^{N_l} \mathbf{x}_j^{(l)} \quad (5.20)$$

donde $\mathbf{x}_j^{(l)}, j = 1, 2, \dots, N_l$ representa los vectores de características pertenecientes a la clase C_l usados como ejemplo. Bajo el modelo de Razonamiento Probabilístico (PMR-1) derivado en [Liu & Wechsler, 1998] las matrices de covarianza son idénticas y se estiman como sigue

$$\boldsymbol{\Sigma}_l = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2\} \quad (5.21)$$

$$\sigma_i^2 = \frac{1}{L} \sum_{l=1}^L \left\{ \frac{1}{N_l - 1} \sum_{j=1}^{N_l} (x_{ji}^{(l)} - m_{li})^2 \right\} \quad (5.22)$$

$x_{ji}^{(l)}$ es el i -ésimo elemento del vector $\mathbf{x}_j^{(l)}$, m_{li} es el i -ésimo elemento de \mathbf{M}_l , y m es el tamaño de los vectores de características.

5.3.4. Análisis Discriminante Lineal

El Análisis Discriminante Lineal (LDA, del inglés Linear Discriminant Analysis) es un modelo de clasificación lineal similar al clasificador de Fisher. Consiste en proyectar los datos de entrada sobre una dimensión y asignar las etiquetas en base a un umbral. Consideremos el caso de clasificación binaria y supongamos que se toma un vector de características n -dimensional, \mathbf{x} , y se proyecta en una dimensión usando

$$y = \mathbf{w}^T \mathbf{x} \quad (5.23)$$

Podemos obtener el clasificador estableciendo un umbral sobre y que clasifique \mathbf{x} como perteneciente a la clase C_1 si $y \geq -w_0$ y como perteneciente a la clase C_2 en caso contrario. En general, la proyección en una sola dimensión conduce a una considerable pérdida de información y clases que se pueden separar correctamente, en el espacio n -dimensional, pueden llegar a estar fuertemente solapadas. Sin embargo, mediante el ajuste de las componentes del vector \mathbf{w} se puede seleccionar una proyección que maximice la separación de clases. Consideremos el caso de dos clases en las que existan N_1 objetos en la clase C_1 y N_2 objetos en la clase C_2 de modo que los vectores medios de ambas clases vienen dados por

$$\mu_1 = \frac{1}{N_1} \sum_{n \in C_1} \mathbf{x}_n, \mu_2 = \frac{1}{N_2} \sum_{n \in C_2} \mathbf{x}_n \quad (5.24)$$

La medida más simple de separación de clases cuando se proyecta sobre \mathbf{w} es la separación entre las proyecciones de las medias. Esto sugiere elegir \mathbf{w} de modo que maxice:

$$\mu_1 - \mu_2 = \mathbf{w}^T (\mu_1 - \mu_2) \quad (5.25)$$

donde $\mu_k = \mathbf{W}^T \mu_k$ es la media de los datos proyectados de la clase C_k .

5.3.5. Clasificador de los k Vecinos más Cercanos

El algoritmo de los k Vecinos más Cercanos (KNN, del inglés k Nearest Neighbor) fue diseñado por Fix y Hodges en 1951 y es uno de los algoritmos de

clasificación más simples que existen [Shakhnarovich *et al.*, 2006]. Se basa en la idea de cercanía de los distintos objetos que queremos clasificar. A grandes rasgos, para clasificar un objeto, KNN busca los k ejemplos más similares (entre todos los ejemplos que el algoritmo almacenó durante la fase de entrenamiento) y asigna a ese objeto la etiqueta mayoritaria entre los ejemplos buscados. La fase de entrenamiento consiste en almacenar los ejemplos con sus etiquetas.

Formalmente, el clasificador KNN se define como

$$\hat{f}(\mathbf{x}') = \operatorname{argmax}_{v \in V} \sum_{i=1}^k \delta(v, f(\mathbf{x}_i)) \quad (5.26)$$

donde \mathbf{x}' es el ejemplo que queremos clasificar y cuya clase se desconoce, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ son los k vecinos más cercanos a \mathbf{x}' y V es el conjunto de todas las etiquetas. La función $\delta(x, y)$ viene dada por

$$\delta(x, y) = \begin{cases} 1, & \mathbf{x} = y \\ 0, & \mathbf{x} \neq y \end{cases} \quad (5.27)$$

Uno de los problemas que presenta el algoritmo es la dificultad para elegir el parámetro k adecuado, valores grandes de k crean límites entre clases parecidas mientras que con valores pequeños de k el algoritmo se ve muy afectado por el ruido. La figura 5.1 muestra un ejemplo de funcionamiento de este algoritmo en el que se desea clasificar el círculo negro central. Para $k = 3$ éste es clasificado con la clase triángulo, ya que hay sólo un cuadrado y 2 triángulos, dentro del círculo que los contiene. Si $k = 5$, el círculo es clasificado con la clase cuadrado, ya que hay 2 triángulos y 3 cuadrados, dentro del círculo externo.

Existen variantes de este algoritmo que asignan pesos a los atributos que componen los vectores de características para dar más importancia a aquellos atributos más útiles para clasificar evitando así que los atributos irrelevantes (que por lo general son la mayoría) determinen la clasificación. De manera similar, otras variantes del algoritmo tienen en cuenta la distancia a los k vecinos más cercanos para determinar la clase. En el ejemplo de la figura 5.1, para $k = 5$ hay mayoría de cuadrados (los vecinos son 2 triángulos y 3 cuadrados) sin embargo los dos triángulos están más cerca del objeto que queremos clasificar por lo que esta variante le asignaría la clase triángulo.

5.3.6. Árboles de Decisión

Los Árboles de Decisión (DT, del inglés Decision Trees) [Breiman *et al.*, 1984a] suponen un método similar a los sistemas de predicción basados en re-

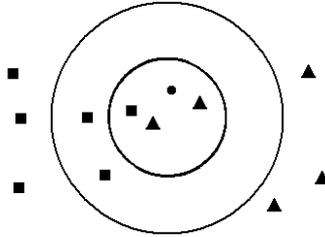


Figura 5.1: Ejemplo de funcionamiento del algoritmo KNN

glas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva. Los clasificadores de árboles binarios estructurados son construidos por repetidas particiones de un conjunto en dos subconjuntos. Estas particiones se hacen estableciendo una expresión booleana de forma que queden en un subconjunto los datos para los que esa expresión es verdadera y en el otro subconjunto los datos para los que la expresión es falsa. La figura 5.2 muestra un ejemplo de funcionamiento de un clasificador de este tipo.

Una vez construido el árbol, para clasificar un objeto, dado su vector de características, se parte del nodo raíz y se van evaluando las condiciones sobre el objeto en cuestión, desplazándose por los nodos hasta llegar a un nodo hoja que nos dará la clase del objeto.

5.3.7. Random Forests

Una manera de incrementar la precisión que se consigue con un clasificador es agrupar varios clasificadores. Generalmente los clasificadores que se agrupan suelen ser del mismo tipo, diferenciándose unos de otros dentro de la agrupación en el conjunto de datos de entrenamiento o en los parámetros que usan. No obstante, también existen enfoques que combinan clasificadores generados por distintos métodos, como es el caso de Stacking [Wolpert, 1992].

Uno de los métodos más populares para crear estas agrupaciones de clasificadores es el conocido como Boosting [Schapire, 2001]. Este término engloba toda una familia de métodos de la que AdaBoost es la variante más conocida. Estos métodos trabajan asignando un peso a cada ejemplo. Inicialmente todos

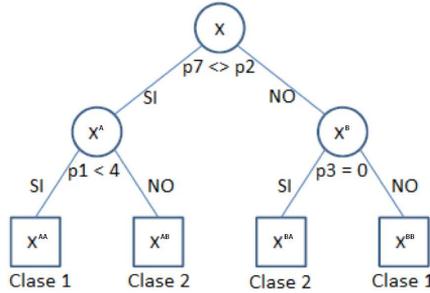


Figura 5.2: Árbol de decisión. Inicialmente, los datos de ejemplo, $\mathbf{x} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ (con $\mathbf{x}_i = \{p_1, p_2, \dots, p_N\}$), se reparten en dos subconjuntos disjuntos, en \mathbf{x}^A están aquellos \mathbf{x}_i que cumplen $p_7 \neq p_2$ y \mathbf{x}^B los que no. El árbol se construye, repitiendo este proceso con los subconjuntos generados hasta llegar a subconjuntos en los que todos los \mathbf{x}_i sean de la misma clase

los ejemplos tienen el mismo peso. En cada iteración se añade un clasificador al conjunto utilizando algún método de aprendizaje y teniendo en cuenta la distribución de pesos. A continuación el peso de cada ejemplo se reajusta: si el clasificador añadido le asigna la clase correcta, disminuye su peso, sino, aumenta. Así se consigue que los siguientes clasificadores que se añadan se centren en los ejemplos mal clasificados por los otros clasificadores del conjunto [Rodríguez & Maudes, 2008].

En el método denominado Bagging (o Bootstrap aggregating) [Breiman, 1996] se usan clasificadores del mismo tipo y cada clasificador se construye usando un conjunto de datos diferente, que es una muestra con reemplazamiento del conjunto de datos original. Normalmente el tamaño de la muestra coincide con el del conjunto de datos de partida pero los datos difieren ya que en una muestra, una instancia puede seleccionarse varias veces mientras que otra puede no seleccionarse.

El método de Random Forests (RF) [Breiman, 2001] es una variante de Bagging en la que los clasificadores que se combinan son árboles de decisión. Cada árbol se entrena con un subconjunto de N ejemplos seleccionados aleatoriamente y con reemplazamiento de un conjunto de N ejemplos. Por tanto, el tamaño de cada subconjunto de entrenamiento es igual al tamaño del conjunto de datos de partida. Si N es suficientemente grande se espera que cada subconjunto tenga un 63% de ejemplos únicos, el resto serán duplicados. Para clasificar un nuevo ejemplo dado por su vector de características, se presenta el vector de

características a cada árbol de decisión. Cada uno de ellos emite un voto y se considera la clase que obtenga un mayor número de votos.

Definición. Sea Θ_k un conjunto de entrenamiento de entre $\Theta_1, \dots, \Theta_n$ conjuntos de entrenamiento generados independientemente. Un clasificador Random Forest consiste en un conjunto de n árboles de decisión $\{h(\mathbf{x}, \Theta_k)\}$, $k = 1, \dots, n$ donde los conjuntos de entrenamiento $\{\Theta_k\}$ son independientes e idénticamente distribuidos y cada árbol emite un voto sobre la clase del objeto de entrada, \mathbf{x} . La decisión final sobre la clase de \mathbf{x} se toma por mayoría.

Este método es poco sensible al ruido y al sobreentrenamiento, ya que el muestreo no se basa en la ponderación. Además, es computacionalmente más eficiente que los métodos basados Boosting.

Convergencia

Dado un conjunto de clasificadores $h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_K(\mathbf{x})$, con un conjunto de entrenamiento obtenido a partir de una distribución aleatoria Y, \mathbf{X} , se define la función margen como

$$mg(\mathbf{X}, Y) = av_k I(h_k(\mathbf{X}) = Y) - \max_{j \neq Y} av_k I(h_k(\mathbf{X}) = j) \quad (5.28)$$

donde I es la función indicatriz ($I(e)$ vale 1 si la expresión e es verdadera y 0 en otro caso) y $av_k X$ representa la media X para todo k . La función margen, mg , mide la amplitud para la que el número medio de votos para la clase correcta en \mathbf{X}, Y excede el número medio de votos para cualquier otra clase. Cuanto mayor sea el margen más confianza podremos tener en el resultado de la clasificación. El error de generalización se define como

$$PE^* = P_{\mathbf{X}, Y}(mg(\mathbf{X}, Y) < 0) \quad (5.29)$$

donde los subíndices \mathbf{X}, Y indican que la probabilidad es sobre el espacio \mathbf{X}, Y . En Random Forests, $h_k(\mathbf{X}) = h(\mathbf{X}, \Theta_k)$. Para un número grande de árboles sigue la Ley Fuerte de los Números Grandes:

Teorema. Cuando el número de árboles crece, para casi todas las secuencias Θ_1, \dots, PE^* , el algoritmo converge a

$$P_{\mathbf{X}, Y}(P_{\Theta}(h(\mathbf{X}, \Theta) = Y) - \max_{j \neq Y} P_{\Theta}(h(\mathbf{X}, \Theta) = j) < 0) \quad (5.30)$$

Este resultado explica por qué Random Forests no sufre sobreentrenamiento cuando hay muchos árboles sino que se alcanza un valor límite del error de generalización.

Fuerza y correlación

Con Random Forests se puede definir una cota superior para el error de generalización en términos de dos parámetros: a) una medida de cómo de precisos son los clasificadores individuales y b) una medida de la dependencia entre ellos. La interacción entre estos dos parámetros proporciona los fundamentos para comprender el funcionamiento de Random Forests.

Definición. La función margen para un clasificador Random Forest es

$$mr(\mathbf{X}, Y) = P_{\Theta}(h(\mathbf{X}, \Theta) = Y) - \max_{j \neq Y} P_{\Theta}(h(\mathbf{X}, \Theta) = j) \quad (5.31)$$

y la fuerza del conjunto de clasificadores $\{h(\mathbf{x}, \Theta)\}$ es

$$s = E_{\mathbf{X}, Y} mr(\mathbf{X}, Y) \quad (5.32)$$

Asumiendo que $s \geq 0$, la desigualdad de Chebychev da

$$PE^* \leq \frac{\text{var}(mr)}{s^2} \quad (5.33)$$

Una expresión más reveladora para la varianza de mr se deriva de lo siguiente. Sea

$$\hat{j}(\mathbf{X}, Y) = \arg \max_{j \neq Y} P_{\Theta}(h(\mathbf{X}, \Theta) = j) \quad (5.34)$$

así,

$$\begin{aligned} mr(\mathbf{X}, Y) &= P_{\Theta}(h(\mathbf{X}, \Theta) = Y) - P_{\Theta}(h(\mathbf{X}, \Theta) = \hat{j}(\mathbf{X}, Y)) \\ &= E_{\Theta}[I(h(\mathbf{X}, \Theta) = Y) - I(h(\mathbf{X}, \Theta) = \hat{j}(\mathbf{X}, Y))] \end{aligned}$$

Definición. La función de margen bruto es

$$rmg(\Theta, \mathbf{X}, Y) = I(h(\mathbf{X}, \Theta) = Y) - I(h(\mathbf{X}, \Theta) = \hat{j}(\mathbf{X}, Y)) \quad (5.35)$$

De este modo, $mr(\mathbf{X}, Y)$ es el valor esperado de $rmg(\Theta, \mathbf{X}, Y)$ con respecto a Θ . Para cualquier función de identidad f ,

$$[E_{\Theta} f(\Theta)]^2 = E_{\Theta, \Theta'} f(\Theta) f(\Theta') \quad (5.36)$$

donde Θ y Θ' son independientes e idénticamente distribuidos. Esto implica que

$$mr(\mathbf{X}, Y)^2 = E_{\Theta, \Theta'} rmg(\Theta, \mathbf{X}, Y) rmg(\Theta', \mathbf{X}, Y) \quad (5.37)$$

Usando 5.37 obtenemos

$$\begin{aligned} \text{var}(mr) &= E_{\Theta, \Theta'} (\text{cov}_{\mathbf{X}, Y} rmg(\Theta, \mathbf{X}, Y) rmg(\Theta', \mathbf{X}, Y)) \\ &= E_{\Theta, \Theta'} (\rho(\Theta, \Theta') sd(\Theta) sd(\Theta')) \end{aligned} \quad (5.38)$$

donde $\rho(\Theta, \Theta')$ es la correlación entre $rmg(\Theta, \mathbf{X}, Y)$ y $rmg(\Theta', \mathbf{X}, Y)$ manteniendo Θ y Θ' fijos y $sd(\Theta)$ es la desviación estándar de $rmg(\Theta, \mathbf{X}, Y)$ manteniendo Θ fijo. Por tanto,

$$\begin{aligned} \text{var}(mr) &= \bar{\rho} (E_{\Theta} sd(\Theta))^2 \\ &\leq \bar{\rho} E_{\Theta} \text{var}(\Theta) \end{aligned} \quad (5.39)$$

donde $\bar{\rho}$ es el valor medio de la correlación, es decir,

$$\bar{\rho} = \frac{E_{\Theta, \Theta'} (\rho(\Theta, \Theta') sd(\Theta) sd(\Theta'))}{E_{\Theta, \Theta'} (sd(\Theta) sd(\Theta'))} \quad (5.40)$$

Teniendo en cuenta que

$$\begin{aligned} E_{\Theta} \text{var}(\Theta) &\leq E_{\Theta} (E_{\mathbf{X}, Y} rmg(\Theta, \mathbf{X}, Y))^2 - s^2 \\ &\leq 1 - s^2 \end{aligned} \quad (5.41)$$

junto con las ecuaciones (5.32) y (5.39) se obtiene:

Teorema. Una cota superior para el error de generalización viene dada por

$$PE^* \leq \frac{\bar{\rho}(1 - s^2)}{s^2} \quad (5.42)$$

Las ecuaciones anteriores muestran que los dos valores involucrados en el error de generalización para Random Forests son la fuerza de los clasificadores individuales y la correlación entre ellos, en términos de la función de margen bruto. La razón c/s^2 es la correlación dividida por el cuadrado de la fuerza. Para comprender el funcionamiento de Random Forests, este valor es de gran ayuda. Cuando menor sea, mejor.

Definición. La razón c/s^2 para un clasificador Random Forest se define como

$$c/s^2 = \frac{\bar{\rho}}{s^2} \quad (5.43)$$

En caso de que haya solamente dos clases, la función de margen queda así:

$$mr(\mathbf{X}, Y) = 2P_{\Theta}(h(\mathbf{X}, \Theta) = Y) - 1 \quad (5.44)$$

El requerimiento de que la fuerza sea positiva (véase ecuación 5.33) es similar a la condición $E_{\mathbf{X}, Y} P_{\Theta}(h(\mathbf{X}, \Theta) = Y) > .5$. La función de margen bruto es $2I(h(\mathbf{X}, \Theta) = Y) - 1$ y la correlación, $\bar{\rho}$, está entre $I(h(\mathbf{X}, \Theta) = Y)$ y $I(h(\mathbf{X}, \Theta') = Y)$. En particular si los valores de Y están entre -1 y 1 ,

$$\bar{\rho} = E_{\Theta, \Theta'}[\rho(h(\cdot, \Theta), h(\cdot, \Theta'))] \quad (5.45)$$

de forma que $\bar{\rho}$ es la correlación entre 2 árboles diferentes del conjunto, promediada sobre la distribución Θ, Θ' .

5.3.8. Máquinas de Vectores Soporte

Desde su introducción a finales de los años 70 [Vapnik, 1982], las Máquinas de Vectores de Soporte (SVM, del inglés Support Vector Machines) marcaron el comienzo de una nueva era en el paradigma del aprendizaje a partir de ejemplos [Burges, 1998; Joachims, 1998]. SVM ha atraído recientemente la atención de la comunidad dedicada al reconocimiento de patrones debido a la cantidad de méritos derivados de la Teoría del Aprendizaje Estadístico [Vapnik, 1995, 1998] desarrollada por Vladimir Vapnik en AT&T. Distinguiremos 3 casos: SVM lineal con clases linealmente separables, SVM lineal con clases linealmente no separables y SVM no lineal.

SVM lineal cuando las clases son linealmente separables

Los clasificadores lineales definen hipersuperficies o hiperplanos de decisión en espacios multidimensionales, esto es:

$$g(\mathbf{x}) = \mathbf{w}^T(\mathbf{x}) + w_0 = 0 \quad (5.46)$$

donde \mathbf{w} se conoce como vector de peso y w_0 como el umbral. El vector de pesos \mathbf{w} es ortogonal al hiperplano de decisión y la tarea de optimización consiste en encontrar el conjunto de parámetros w_i , $i = 1, \dots, N$, que definen el hiperplano de decisión.

Sean \mathbf{x}_i , $i = 1, 2, \dots, l$, los vectores de características del conjunto de entrenamiento, \mathbf{X} . Éstos pertenecen a una de las dos clases C_1 o C_2 , las cuales se asumen linealmente separables. El objetivo es diseñar el hiperplano, $g(\mathbf{x})$, que clasifica correctamente todos los vectores de entrenamiento y que dista lo mismo de los puntos más cercanos de C_1 y C_2 . La selección se hace de esta forma para favorecer la generalización, esto es, la capacidad del clasificador para clasificar satisfactoriamente vectores diferentes a los del conjunto de entrenamiento. Puesto que la distancia desde un punto, \mathbf{x} , al hiperplano viene dada por:

$$z = \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|} \quad (5.47)$$

escalamos \mathbf{w} y w_0 de modo que el valor de $g(\mathbf{x})$ en los puntos más cercanos sea +1 para el punto más cercano en C_1 y -1 para el punto más cercano en C_2 . Esto reduce el problema de optimización a maximizar el margen:

$$\frac{1}{\|\mathbf{w}\|} + \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (5.48)$$

sujeto a las condiciones:

$$\begin{aligned} \mathbf{w}^T \mathbf{x} + w_0 &\geq 1, & \forall \mathbf{x} \in C_1, \\ \mathbf{w}^T \mathbf{x} + w_0 &\leq -1, & \forall \mathbf{x} \in C_2, \end{aligned} \quad (5.49)$$

Para cada \mathbf{x}_i denotamos el correspondiente indicador de clase y_i (+1 para C_1 y -1 para C_2), por tanto nuestro objetivo se reduce a calcular los parámetros \mathbf{w} y w_0 del hiperplano de manera que se minimice la siguiente expresión:

$$\mathbf{J}(\mathbf{w}) \equiv \frac{1}{2} \|\mathbf{w}\|^2 \quad (5.50)$$

sujeto a

$$y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1 \quad \text{para } i = 1, 2, \dots, N \quad (5.51)$$

Obviamente, minimizando la norma el margen se hace mínimo. Esto es una tarea de optimización cuadrática (no lineal) sujeta a un conjunto de restricciones (inecuaciones lineales). De acuerdo a Karush-Kuhn-Tucker (KKT) debe cumplirse:

$$\frac{\partial}{\partial \mathbf{w}} \mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \mathbf{0} \quad (5.52)$$

$$\frac{\partial}{\partial w_0} \mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\lambda}) = 0 \quad (5.53)$$

$$\lambda_i \geq 0 \quad \text{para } i = 1, 2, \dots, N \quad (5.54)$$

$$\lambda_i [y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1] = 0 \quad \text{para } i = 1, 2, \dots, N \quad (5.55)$$

donde λ es el vector de multiplicadores de Lagrange, λ_i , y $\mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\lambda})$ es la función Lagrangiana definida como

$$\mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \lambda_i [y_i(\mathbf{w}^T \mathbf{x}_i + w_0) - 1] \quad (5.56)$$

Combinando 5.56 con 5.52 y 5.53 resulta

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i \quad (5.57)$$

$$\sum_{i=1}^N \lambda_i y_i = 0 \quad (5.58)$$

Los multiplicadores de Lagrange pueden ser cero o positivos. Por tanto, el vector de parámetros \mathbf{w} de la solución óptima es una combinación lineal de $N_s \leq N$ vectores característicos asociados a $\lambda_i \neq 0$, es decir,

$$\mathbf{w} = \sum_{i=1}^{N_s} \lambda_i y_i \mathbf{x}_i \quad (5.59)$$

Éstos se conocen como los *vectores de soporte* y el hiperplano clasificador óptimo como *máquina de vectores de soporte* (SVM). Al igual que para el conjunto de restricciones en 5.55 para $\lambda_i \neq 0$, los vectores de soporte caen en uno de los dos hiperplanos

$$\mathbf{w}^T \mathbf{x} + w_0 = \pm 1 \quad (5.60)$$

es decir, son los vectores de entrenamiento que están más cerca del clasificador lineal, y constituyen los elementos críticos del conjunto de entrenamiento.

Mientras que \mathbf{w} se da explícitamente, w_0 se puede obtener mediante una de las condiciones 5.55. En la práctica, w_0 se calcula como un valor medio obtenido usando todas las condiciones de este tipo. Por otro lado, las propiedades de la función de coste 5.50 garantizan que la matriz Hessiana correspondiente es definida positiva. Además, las restricciones consisten en funciones lineales. Estas dos condiciones garantizan que cualquier mínimo local es también global y único. El hiperplano clasificador de una máquina de vectores de soporte es único.

El cálculo de los parámetros involucrados no siempre es una tarea fácil. Se trata de un problema de la familia de programación convexa que se resuelve considerando la denominada *dualidad Lagrangiana*. El problema puede formularse equivalentemente como maximizar

$$\mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\lambda}) \quad (5.61)$$

sujeto a

$$\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i \quad (5.62)$$

$$\sum_{i=1}^N \lambda_i y_i = 0 \quad (5.63)$$

$$\boldsymbol{\lambda} \geq \mathbf{0} \quad (5.64)$$

Las dos restricciones de igualdad son el resultado de igualar a cero el gradiente de la Lagrangiana, con respecto a \mathbf{w} y w_0 . Los vectores de características de entrenamiento aparecen en el problema mediante las restricciones de igualdad y no mediante las inecuaciones, lo cual hace que el problema sea más fácil

de manejar. Sustituyendo 5.62 y 5.63 en 5.61 y haciendo algunas operaciones llegamos a una tarea de optimización equivalente:

$$\max_{\lambda} \left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right) \quad (5.65)$$

sujeto a

$$\sum_{i=1}^N \lambda_i y_i = 0 \quad (5.66)$$

$$\boldsymbol{\lambda} \geq \mathbf{0} \quad (5.67)$$

Una vez que los multiplicadores de Lagrange han sido calculados maximizando 5.65, el hiperplano óptimo se calcula vía 5.62 y w_0 como antes. Lo más interesante de este enfoque es que la función coste no depende explícitamente de la dimensionalidad del espacio de entrada, lo cual permite generalizaciones eficientes para el caso de clases linealmente no separables.

Por otro lado, aunque el hiperplano óptimo resultante es único, no existe garantía de la unicidad de los multiplicadores de Lagrange asociados, λ_i , es decir, la expansión de \mathbf{w} en términos de vectores soporte en 5.62 puede no ser única, aunque el resultado final es único.

SVM lineal cuando las clases no son linealmente separables

En el caso de que las clases no sean separables, lo dicho anteriormente deja de ser válido. Cualquier intento de dibujar un hiperplano no conseguirá una banda de separación de clases sin puntos dentro de ella. Recordemos que el margen se define como la distancia entre el par de hiperplanos paralelo descritos por:

$$\mathbf{w}^T \mathbf{x} + w_0 = \pm 1 \quad (5.68)$$

Los vectores de características de entrenamiento ahora pertenecen a una de las siguientes tres categorías

- Vectores que caen fuera de la banda y que son correctamente clasificados. Estos vectores cumplen con las restricciones en 5.50.
- Vectores que caen dentro de la banda y que son correctamente clasificados. Estos puntos satisfacen la inecuación

$$0 \leq y_i (\mathbf{w}^T \mathbf{x} + w_0) < 1 \quad (5.69)$$

- Vectores que son clasificados erróneamente. Éstos están rodeados por círculos y cumplen la inecuación

$$y_i(\mathbf{w}^T \mathbf{x} + w_0) < 0 \quad (5.70)$$

Estos tres casos se pueden tratar como un solo tipo de restricciones introduciendo un nuevo conjunto de variables

$$y_i(\mathbf{w}^T \mathbf{x} + w_0) \geq 1 - \xi_i \quad (5.71)$$

La primera categoría de datos corresponde con $\xi_i = 0$, la segunda con $0 < \xi_i \leq 1$ y la tercera con $\xi_i > 1$. Las variables ξ_i se conocen como *variables débiles*. La tarea de optimización se vuelve más complicada, aunque se basa en los mismos principios que antes. El objetivo ahora es hacer el margen tan grande como sea posible pero al mismo tiempo mantener la cantidad de puntos con $\xi_i \geq 0$ tan pequeña como sea posible. En términos matemáticos, esto es equivalente a minimizar la función de coste

$$J(\mathbf{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \mathbf{I}(\xi_i), \quad (5.72)$$

donde $\boldsymbol{\xi}$ es el vector de parámetros ξ_i y

$$\mathbf{I}(\xi_i) = \begin{cases} 1, & \xi_i > 0 \\ 0, & \xi_i = 0 \end{cases} \quad (5.73)$$

El parámetro C es una constante positiva que controla la influencia relativa de los dos términos competitivos. Sin embargo, la optimización de arriba es difícil puesto que incluye una función discontinua $\mathbf{I}(\cdot)$. Como es común en casos así, se elige optimizar una función de coste estrechamente relacionada, y el objetivo se convierte en minimizar

$$J(\mathbf{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i, \quad (5.74)$$

sujeto a

$$\begin{aligned} y_i(\mathbf{w}^T \mathbf{x}_i + w_0) &\geq 1 - \xi_i, & i = 1, 2, \dots, N \\ \xi_i &> 0, & i = 1, 2, \dots, N \end{aligned} \quad (5.75)$$

El problema es de nuevo un problema de programación convexa, y la Lagrangiana correspondiente viene dada por

$$\mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \mu_i \xi_i - \sum_{i=1}^N \lambda_i [y_i (\mathbf{w}^T \mathbf{x}_i + w_0) - 1 + \xi_i] \quad (5.76)$$

Realizando pasos similares a los del caso de clases separables, llegamos al siguiente problema de optimización equivalente

$$\max_{\boldsymbol{\lambda}} \left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right) \quad (5.77)$$

sujeto a

$$0 \leq \lambda_i \leq C, \text{ para } i = 1, 2, \dots, N \quad (5.78)$$

$$\sum_{i=1}^N \lambda_i y_i = 0 \quad (5.79)$$

La única diferencia con el caso previamente considerado de clases linealmente separables está en la primera de las dos restricciones, donde es necesario limitar por arriba a los multiplicadores de Lagrange por C . El caso linealmente separable corresponde con $C \rightarrow \infty$. Las variables débiles ξ_i , y sus multiplicadores de Lagrange asociados, μ_i , no intervienen en el problema explícitamente. Su presencia está reflejada indirectamente mediante C .

En todo este desarrollo se ha considerado sólo el caso de clasificación con dos clases. En el caso de M -clases, se puede extender fácilmente mirando el problema como M problemas de dos clases. Para cada una de las clases, tratamos de diseñar una función discriminante óptima, $g_i(\mathbf{x})$, $i=1,2,\dots,M$, de modo que $g_i(\mathbf{x}) > g_j(\mathbf{x}), \forall j \neq i$ si $\mathbf{x} \in C_i$. Adoptando la metodología de SVM podemos diseñar las funciones discriminantes de modo que $g_i(\mathbf{x}) = 0$ sea el hiperplano óptimo para separar la clase C_i de todas las demás, dado por supuesto que esto es posible. Así, la función lineal resultante dará $g_i(\mathbf{x}) > 0$ para $\mathbf{x} \in C_i$ y $g_i(\mathbf{x}) < 0$ en caso contrario. La clasificación se consigue de acuerdo a la siguiente regla

$$\text{Asignar } \mathbf{x} \text{ a } C_i \text{ si } i = \arg_k \max \{g_k(\mathbf{x})\}$$

Esta técnica, sin embargo, puede conducir a regiones indeterminadas, donde más de un $g_i(\mathbf{x})$ es positivo. Otra aproximación es extender la formulación matemática de SVM de dos clases al problema de M clases.

SVM no lineal

En el apartado anterior se discutieron las máquinas de vectores de soporte (SVM) como una metodología óptima de diseño de un clasificador lineal. Asumimos ahora que existe un mapeo

$$\mathbf{x} \in \mathbb{R}^l \rightarrow \mathbf{y} \in \mathbb{R}^k$$

desde el espacio de entrada a un espacio k -dimensional, donde las clases se pueden separar satisfactoriamente por un hiperplano lineal. Recordamos que los vectores de características participan por pares mediante la operación del producto interno. También, una vez que el hiperplano óptimo (\mathbf{w}, w_0) se ha calculado, la clasificación se realiza en función del signo de

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 = \sum_{i=0}^{N_s} \lambda_i y_i \mathbf{x}_i^T \mathbf{x} + w_0 \quad (5.80)$$

donde N_s es el número de vectores de soporte. Así, una vez más, sólo el producto interno entra en escena. Si el diseño se va a llevar a cabo en el espacio k -dimensional, la única diferencia es que los vectores involucrados estarán en los mapeos k -dimensionales del vector de características original. Una simple ojeada a esto nos llevaría a la conclusión de que ahora la complejidad es mucho mayor, puesto que, habitualmente, k es mucho más alto que la dimensión l del espacio de entrada, para poder hacer las clases linealmente separables. Sin embargo, vamos a ver con un ejemplo que se cumple una propiedad interesante. Supongamos que

$$\mathbf{x} \in \mathbb{R}^2 \rightarrow \mathbf{y} = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix} \quad (5.81)$$

Con un poco de álgebra sencilla, se puede mostrar que

$$\mathbf{y}_i^T \mathbf{y}_j = (\mathbf{x}_i^T \mathbf{x}_j)^2 \quad (5.82)$$

o dicho de otro modo, el producto interno de los vectores en el nuevo (y mayor) espacio dimensional se puede expresar como función del producto interno de los correspondientes vectores en el espacio de características original.

Una vez que se ha adoptado el *kernel* adecuado, que implícitamente define un mapeo a un espacio de dimensión mayor, la tarea de clasificación se convierte en

$$\max_{\lambda} \left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \quad (5.83)$$

sujeto a

$$0 \leq \lambda_i \leq C, \quad i = 1, 2, \dots, N \quad (5.84)$$

$$\sum_i \lambda_i y_i = 0 \quad (5.85)$$

y el clasificador lineal resultante es

$$\begin{aligned} & \text{Asignar } \mathbf{x} \text{ a } C_1 \text{ si } g(\mathbf{x}) > 0 \text{ y a } C_2 \text{ si } g(\mathbf{x}) < 0 \text{ siendo} \\ & g(\mathbf{x}) = \sum_{i=1}^{N_s} \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + w_0 \end{aligned}$$

La figura 5.3 muestra la arquitectura correspondiente. El número de nodos viene determinado por el número de vectores de soporte N_s . Los nodos realizan el producto interno entre el mapeo de \mathbf{x} y los correspondientes mapeos de los vectores de soporte en el espacio de dimensión mayor, mediante la operación de *kernel*.

Observaciones:

- Una característica notable de las máquinas de vectores de soporte es que la complejidad computacional es independiente de la dimensionalidad del espacio kernel, donde las características de entrada son mapeadas. Así, se posibilita realizar el diseño en un espacio de dimensión alta sin tener que adoptar modelos explícitos que requieren una gran cantidad de parámetros.
- Una limitación importante de las máquinas de vectores de soporte es la alta carga computacional que se requiere, tanto durante el entrenamiento como durante el test. Para problemas con una cantidad relativamente pequeña de datos de entrenamiento, se puede usar cualquier algoritmo de optimización de propósito general. Sin embargo, para una gran cantidad de puntos de entrenamiento (del orden de unos miles), se requiere un

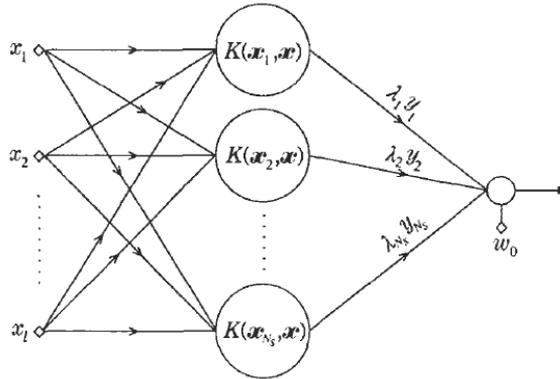


Figura 5.3: La arquitectura de SVM usando las funciones de kernel.

tratamiento especial. Para solventar este problema, se han ideado ciertos procedimientos basados en la descomposición, de una manera o de otra, del problema de optimización en una secuencia de otros más pequeños.

- Otra limitación importante de las máquinas de vectores de soporte es que, hasta ahora, no hay un método práctico para seleccionar la mejor función de kernel. Esto es todavía un problema sin solución.
- Las máquinas de vectores de soporte se han aplicado a una gran cantidad de aplicaciones diversas, que van desde reconocimiento de dígitos manuscritos, el reconocimiento de objetos, identificación de personas y ecualización de canal. Los resultados indican que los clasificadores SVM ponen de manifiesto un comportamiento en general mejorado.

5.4. Extracción de características

En la tarea de clasificación o separación de objetos en clases, tan importante es el método de clasificación escogido como la forma en que construyan los vectores de características que definen a cada objeto. La dimensión de estos vectores debe ser lo suficientemente grande como para recoger todas las propiedades de los objetos teniendo en cuenta que, para mantener el rendimiento, el número de

muestras usadas en el entrenamiento debe crecer proporcionalmente al número de características. La relación entre el número de muestras del entrenamiento, n , y la dimensión de los vectores de características, m , es tratada por la teoría de aprendizaje estadístico.

Consideremos n puntos en un espacio de características m -dimensional. Se asume que los puntos están *bien distribuidos*, de manera que no existe ningún subconjunto de $n-1$ puntos que se sitúen en un hiperplano de dimensión menor $m-1$. El número $O(l, N)$ de agrupaciones que pueden ser formadas por hiperplanos $(m-1)$ -dimensionales para separar los n puntos en dos clases viene dado por [Cover, 1965]:

$$O(n, m) = 2 \sum_{i=0}^m \binom{n-1}{i} \quad (5.86)$$

donde

$$\binom{n-1}{i} = \frac{(n-1)!}{(n-1-i)!i!} \quad (5.87)$$

Por lo tanto, la probabilidad de agrupar n puntos en un espacio de características m -dimensional en 2 clases linealmente separables será:

$$P_n^m = \frac{O(n, m)}{2^n} = \begin{cases} \frac{1}{2^{n-1}} \sum_{i=0}^m \binom{n-1}{i} & n > m + 1 \\ 1 & n \leq m + 1 \end{cases} \quad (5.88)$$

La figura 5.4 muestra la probabilidad P_n^m como una función de m/n . Para espacios de características de baja dimensión ($m/n < 0,3$), la probabilidad de separación de clases P_n^m es prácticamente cero, esto es, los clasificadores lineales tienen un bajo rendimiento discerniendo entre dos clases. Sin embargo, cuando se aumenta la dimensión del espacio de características, la probabilidad de que el conjunto de n puntos sea separable se aproxima a la unidad. Así, como intuitivamente parece claro, añadir información al vector de características mejora la separabilidad de clases para el caso de clasificador lineal.

Por otro lado, el hecho de que un cociente m/n pequeño produce bajos valores para probabilidad de separabilidad puede resolverse para un número fijo n de puntos a clasificar mediante un mapeo a un espacio de dimensión mayor. Esto puede conseguirse a través del uso de *kernels* (vease sección 5.2), que involucran el uso de un producto interno no lineal, convirtiendo el espacio de características en otro de dimensión mayor donde P_n^m aumente, y un clasificador lineal pueda operar satisfactoriamente.

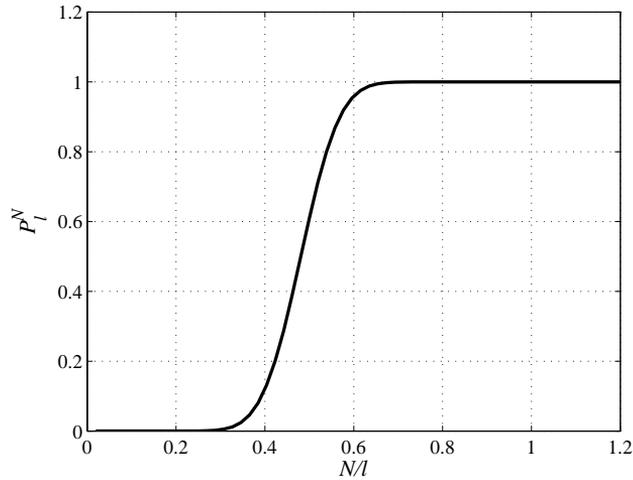


Figura 5.4: Probabilidad de agrupar n puntos en un espacio de características m -dimensional en 2 clases linealmente separables.

A la vista de la figura 5.4 y de la ecuación 5.88 es natural pensar que un aumento de la dimensión del espacio de características será siempre beneficioso para discriminar entre dos clases. Paradójicamente, ocurre exactamente lo contrario, dando lugar al problema del pequeño tamaño muestral (también conocido como la maldición de la dimensionalidad o el fenómeno del máximo). Este problema se describe como una reducción de la eficacia de un clasificador al añadir nuevas características a los vectores de entrenamiento cuando el número de estos es relativamente pequeño en comparación con número de características. El problema radica en que para definir un clasificador en un espacio de características de alta dimensionalidad es necesario estimar un número de parámetros comparable a la dimensión del espacio. Por ejemplo, en el caso de un clasificador lineal, será necesario estimar $m + 1$ parámetros en un espacio de características m -dimensional (ver (5.13)). Por lo tanto, aunque el clasificador separe los datos de entrenamiento satisfactoriamente, la fiabilidad en la estimación de los parámetros del clasificador será baja, ya que se estimarán muchos parámetros con un número muy reducido de vectores de entrenamiento. El clasificador construido con esta limitación tendrá, por consiguiente, una baja capacidad de generalización.

Tabla 5.1: Posibles resultados del test en función de la etiqueta

		Etiqueta		
		Positiva	Negativa	
Test	Positivo	PV	PF	→ <i>Valor predictivo positivo</i>
	Negativo	NF	NV	→ <i>Valor predictivo negativo</i>
		↓	↓	
		<i>Sensibilidad</i>	<i>Especificidad</i>	

El problema del pequeño tamaño muestral justifica el uso de técnicas de reducción de la dimensionalidad del espacio de características, cuando el número de características usadas para diseñar el clasificador es mucho mayor que el número de vectores de entrenamiento disponibles. Además, la reducción del número de características elimina la correlación entre las mismas y disminuye el coste computacional de la clasificación por lo que aunque el problema del pequeño tamaño muestral no exista, es recomendable usar un número reducido de características.

5.5. Evaluación de un clasificador

5.5.1. Medidas de rendimiento

El propósito un todo clasificador es asignar correctamente una etiqueta a un objeto definido por un vector de atributos. En el caso de clasificación binaria, en la que los datos están divididos entre etiquetas positivas o negativas, existirán dos posibles errores que el clasificador puede cometer: clasificar como positivo un patrón que en realidad era negativo o viceversa. Estas posibilidades vienen reflejadas en la tabla 5.1, donde se compara el resultado del test con la etiqueta original, dando lugar a PV (Positivo Verdadero) cuando los dos coinciden en valor positivo, PF (Positivo Falso) cuando el test da positivo mientras la etiqueta original era negativa, NF (Negativo Falso) cuando el test da negativo mientras la etiqueta original era positiva y NV (Negativo Verdadero) cuando los dos coinciden en valor negativo.

La sensibilidad se define como la capacidad de un clasificador para detectar

los positivos verdaderos y se expresa como:

$$\text{Sensibilidad} = \frac{\text{número de } PV}{\text{número de } PV + NF} \quad (5.89)$$

de manera que una sensibilidad del 100 % corresponderá a un clasificador que es capaz de clasificar correctamente todos los objetos etiquetados como positivos. Por lo tanto, si un clasificador con alta sensibilidad da un resultado negativo, éste será muy fiable, lo que, en el caso del diagnóstico asistido por computador, puede ser usado para descartar la enfermedad. La sensibilidad esta relacionada con el error de tipo I en inferencia estadística, que consiste en rechazar la hipótesis nula cuando en realidad es cierta.

Por otro lado, la especificidad se define como la capacidad de un clasificador para detectar negativos verdaderos y se expresa como:

$$\text{Especificidad} = \frac{\text{número de } NV}{\text{número de } NV + PF} \quad (5.90)$$

de manera que una especificidad del 100 % corresponderá a un clasificador que es capaz de clasificar correctamente todos los objetos etiquetados como negativos. En el caso del diagnóstico asistido por computador, un clasificador con alta especificidad es muy útil para confirmar la enfermedad, ya que raramente producirá un resultado positivo que en realidad sea falso. La especificidad esta relacionada con el error de tipo II donde se acepta la hipótesis nula cuando en realidad es falsa.

Sin embargo, tanto altos valores de la sensibilidad como de la especificidad no tienen porque corresponder a un clasificador preciso. Se define la precisión como:

$$\text{Precisión} = \frac{\text{número de } PV + NV}{\text{número de } PV + PF + NF + NV} \quad (5.91)$$

Puede ocurrir que un clasificador tenga valores cercanos al 100 % de sensibilidad y cercanos al 0 % de especificidad. Este clasificador no tendrá capacidad de discernir entre las clases, ya que será un clasificador que tome cualquier patrón como positivo. Esto es equivalente a una clasificación al azar, ya que su precisión rondará el 50 % para una muestra sin preponderancia de ninguna de las dos clases. El clasificador deseable será aquel que tenga valores altos de sensibilidad, especificidad y precisión simultáneamente, y no sólo de alguno de ellos por separado.

Otros parámetros que pueden resultar interesantes son los valores predictivos. Éstos hacen referencia a la validez de un resultado de clasificación positivo/negativo (valor predictivo positivo/negativo). Se podrá confiar más en un resultado positivo de un clasificador con un vpp alto que uno con un vpp menor. Sin embargo, los valores predictivos dependen de la preponderancia de las clases, denominada prevalencia, término de epidemiología que determina la proporción de individuos de una población que, en este caso, padece la enfermedad. Si el conjunto de test no tiene igual número de positivos que de negativos, habrán de usarse las fracciones de probabilidad positiva (fpp):

$$\text{fpp} = \frac{\text{sensibilidad}}{1 - \text{especificidad}} \quad (5.92)$$

o negativa (fpn):

$$\text{fpn} = \frac{1 - \text{sensibilidad}}{\text{especificidad}} \quad (5.93)$$

que no dependen de la prevalencia.

5.5.2. Curva ROC

A menudo será interesante valorar cómo se modifica el rendimiento de un clasificador al modificar algún parámetro, ya sea del clasificador o en la definición de algún paso anterior. Para este análisis será útil la representación en el espacio ROC (Receiver Operating Characteristic [Fawcett, 2006]), que no es más que una representación bidimensional de la tasa de positivos verdaderos (sensibilidad) frente a la tasa de falsos positivos (1-especificidad).

En este espacio (véase figura 5.5), la mejor predicción corresponderá con un punto cercano a la esquina superior izquierda (D), representando sensibilidad del 100 % (ningún negativo falso) y especificidad 100 % (ningún positivo falso), que producirá también una precisión del 100 %. El punto D se llama también clasificación perfecta. Una clasificación completamente aleatoria daría un punto C en la línea diagonal (llamada la línea de no-discriminación) de la parte inferior izquierda a la parte superior derecha. Por debajo de esta línea estarían resultados peores que la estimación al azar E. Los puntos A y B representan clasificaciones adecuadas, la primera con mayor sensibilidad que especificidad y la segunda con una especificidad mayor que la sensibilidad.

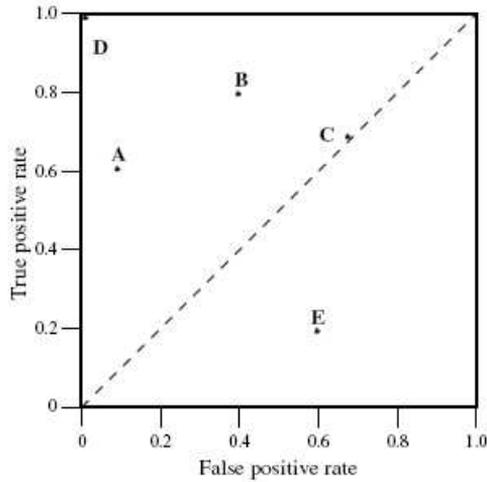


Figura 5.5: Representación en el espacio ROC

5.5.3. Técnicas de validación cruzada

Desarrollada originalmente por Seymour Geisser, la validación cruzada [Kohavi, 1995; Devijver & Kittler, 1982] es una técnica que permite evaluar si el resultado de un análisis estadístico tiene mayor o menor capacidad de generalización sobre un conjunto independiente de datos. Principalmente se emplea en aplicaciones donde el objetivo último es la predicción y se quiere estimar cómo de preciso es el modelo predictivo en la práctica.

Una ronda de validación cruzada consiste en dividir el conjunto muestral en subconjuntos complementarios, y realizar el análisis en un subconjunto (llamado el conjunto de entrenamiento), y posteriormente validar éste análisis en otro subconjunto (llamado subconjunto de validación o de test). Para reducir la variabilidad en la evaluación global de la generalización, se llevan a cabo múltiples rondas de validación cruzada usando distintas particiones, y los resultados de la validación son promediados sobre todas las rondas realizadas. En los métodos de validación cruzada, el conjunto de test no supone un test 'real', ya que la etiqueta de los elementos del conjunto de test es conocida. De esta manera, se puede comparar el resultado del test con la etiqueta original, y determinar si se trata de un PV, PF, NV, o NF. Una vez el proceso de itera sobre cada partición,

se puede calcular cualquiera de las cantidades definidas en la sección anterior.

En los siguientes apartados se describen las principales técnicas de validación cruzada:

Validación por sub-muestreo aleatorio repetido

En cada ronda, la validación por sub-muestreo aleatorio repetido divide de forma aleatoria el conjunto de datos en dos conjuntos. El clasificador es entrenado con uno de los conjuntos y el otro conjunto es usado para validarlo. Tras n rondas, los resultados de la validación son promediados.

La ventaja de este método es que la proporción de la división entrenamiento/validación no depende del número de iteraciones. Por contra, tiene la desventaja de que los conjuntos de validación de las diferentes rondas pueden solaparse y algunas muestras pueden no ser nunca incluidas en el subconjunto de validación mientras que otras pueden ser seleccionadas más de una vez. Este método también exhibe la variación Monte Carlo, esto es, los resultados variarán si el análisis es repetido con diferentes conjuntos aleatorios. Una variante de esta aproximación genera muestras aleatorias de tal forma que el valor de respuesta medio es igual en los subconjuntos de entrenamiento y test. Esto es particularmente útil cuando el conjunto muestral contiene una representación no balanceada en las respuestas de las muestras.

Validación cruzada k -fold

De acuerdo con esta técnica de validación, el conjunto muestral original es dividido en k subconjuntos [Breiman *et al.*, 1984b] de los que uno de ellos se destina a validación y los $k - 1$ restantes se usan como conjunto de entrenamiento. El procedimiento de entrenamiento y validación se repite k veces y en cada una de las cuales se emplea un subconjunto distinto para validación consiguiendo de esta forma que todos los objetos del conjunto muestral sean usados para validar el modelo. Al igual que en otros modelos de validación cruzada, la estimación final es calculada como un promedio de los resultados obtenidos en cada una de las k rondas.

La ventaja de este método sobre el anterior es que todas las observaciones se usan tanto para entrenamiento como para test y cada observación se usa para validación sólo una vez. Este método se suele usar cuando el número de elementos de la muestra es muy grande, o cuando los algoritmos de clasificación son computacionalmente costosos, de manera que se tiene control sobre el número de entrenamientos/validaciones que se realizan a través del parámetro k .

Validación dejar uno fuera

Como el nombre indica, la validación dejar uno fuera consiste en utilizar una sola muestra como observación de test y las restantes observaciones como entrenamiento [Raudys & Jain, 1991]. Este proceso se repite N veces de manera que todas las muestras son usadas una vez como observaciones para la validación, por lo que se considera un caso particular del anterior método de validación en el que $k = N$. La ventaja del método de dejar uno fuera es que el conjunto de entrenamiento es lo más grande que la muestra permite, aumentando la estadística en la estimación de los parámetros del clasificador. Por contra, este método puede ser computacionalmente costoso dado el gran número de veces que se repite el proceso de validación, si el número de elementos de la muestra es grande.

En el caso de diagnóstico médico, éste será a menudo el método de validación elegido, ya que el número de pacientes de una base de datos habitual es suficientemente reducido para que el uso de la validación dejar uno fuera no suponga un coste computacional demasiado elevado.

CAPÍTULO

6

MAPAS ESTADÍSTICOS PARAMÉTRICOS

Los mapas estadísticos paramétricos (de inglés *Statistical Parametric Mapping* o simplemente SPM) [Friston *et al.*, 2007; Gispert *et al.*, 2003] son un conjunto de métodos cuantitativos diseñados para buscar efectos de interés en imágenes funcionales a través de mapas de activación. Por SPM normalmente nos referimos a un software ¹ desarrollado por el centro *Wellcome Trust Centre for Neuroimaging* y que implementa dichos métodos. En este capítulo se muestra una visión general de SPM y su posible aplicación a los sistemas CAD.

¹Puede ser descargado gratuitamente de <http://www.fil.ion.ucl.ac.uk/spm/software/>

6.1. Introducción

El software SPM es una *suite* para MATLAB [Attaway, 2009] que implementa un conjunto de métodos de gran utilidad para el estudio de imágenes funcionales. Desde su primera aparición en 1991, la comunidad de investigadores de neuroimagen funcional ha aceptado y utilizado ampliamente el software SPM gracias a que permite realizar de forma simple una gran cantidad de análisis estadísticos (regresiones, tests t de Student, tests F y análisis de varianza (Anova) incluyendo covariables y permitiendo el modelado de interacciones entre ellas). Actualmente SPM se ha convertido en una referencia en el campo de la neuroimagen y es usado en departamentos de psiquiatría, psicología, neurología, radiología, medicina nuclear, farmacología, ciencias cognitivas y del comportamiento, bioestadística y física biomédica de todo el mundo para la investigación de enfermedades mentales, cuantificación de efectos farmacológicos, estudios cognitivos, realización de análisis longitudinales, estudios intersujeto, e incluso morfométricos [Gispert *et al.*, 2003]. En su versión actual (SPM8) permite el análisis de imágenes fMRI, PET, SPECT, EEG y MEG aunque el campo en el que tiene mayor difusión es en el análisis de resonancias magnéticas funcionales debido a su facilidad para trabajar con secuencias de imágenes de un mismo paciente, algo muy habitual en estudios fMRI.

Los análisis realizados por SPM son univariados ya que los test se realizan vóxel a vóxel (tratando cada vóxel de forma independiente al resto). Por tanto, el resultado de los análisis es un mapa del mismo tamaño que las imágenes funcionales conocido como mapa de estadísticos paramétricos. Estos mapas son

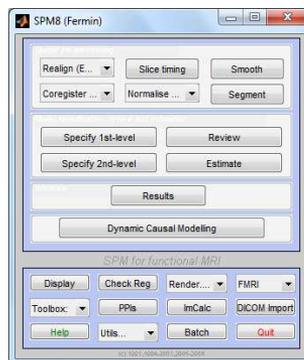


Figura 6.1: Ventana principal de SPM para el análisis de imágenes fMRI

después utilizados para aceptar o rebatir una hipótesis definida *a priori*. Un estudio en SPM consta de tres etapas: preprocesamiento de las imágenes, análisis estadístico e interpretación de los mapas de estadísticos. La etapa de preprocesamiento es imprescindible en todo estudio con imágenes funcionales y solamente debe ser realizada una vez sobre cada imagen. SPM proporciona un amplio conjunto de métodos para realizar el preprocesamiento de las imágenes. En las siguientes secciones se describen las tres etapas en las que se divide un estudio en SPM y al final del capítulo se muestran algunas ideas para el desarrollo de sistemas CAD basados en SPM, si bien, los sistemas multivariados son más apropiados para este propósito.

6.2. Preprocesamiento en SPM

El preprocesamiento es un conjunto de procesos que aseguran que dos imágenes son comparables entre sí, aún cuando éstas proceden de diferentes personas y/o están tomadas con diferentes cámaras. En SPM el preprocesamiento se divide en 4 etapas:

6.2.1. Realineamiento

Esta etapa se aplica solamente a series de imágenes, es decir, cuando hay más de una imagen del mismo paciente (algo muy habitual en imágenes fMRI). El realineamiento trata de eliminar las diferencias en las imágenes debidas a los movimientos del paciente durante el proceso de adquisición. Para ello, realiza únicamente rotaciones y translaciones de las imágenes, tras las cuales, todas las imágenes estarán en un espacio común que normalmente es el de la primera imagen de la serie [Woods *et al.*, 1992].

Aunque en los estudios para el diagnóstico de AD el paciente suele estar relajado y no recibe influencias del ambiente (véase sección 2.3.3), no en todos los estudios de imagen funcional es así. Por ejemplo, en estudios sobre la actividad cerebral de drogodependientes, los paciente suelen estar visionando diferentes fotografías [Zhang *et al.*, 2005]. En estos casos, los posibles movimientos del paciente durante el proceso de adquisición podrían estar relacionados con la tarea que está realizando y es necesario incluirlos como variables en el análisis posterior.

El realineamiento corrige las diferencias relativas a la posición entre distintas imágenes de un mismo paciente pero, si las imágenes proceden de distintos pacientes, habrá diferencias también en cuanto a forma y tamaño del cerebro. Estas

diferencias deben corregirse en la siguiente etapa, la normalización espacial.

6.2.2. Normalización espacial

Una forma de analizar las imágenes cerebrales en SPM consiste en compararlas vóxel a vóxel. Este tipo de análisis requiere que todas las estructuras cerebrales estén en la misma posición anatómica, es decir, que al tomar dos vóxeles cualesquiera de dos imágenes distintas, si ambos tienen las mismas coordenadas, deben pertenecer a la misma estructura cerebral [Salas-Gonzalez *et al.*, 2008a; Ashburner *et al.*, 1997]. Por tanto, antes de realizar un análisis de las imágenes es necesario normalizarlas espacialmente para garantizar dicha correspondencia anatómica.

La normalización espacial consiste en “estirar” las imágenes con el fin de hacerlas encajar con una plantilla. Así, tras este proceso cada región cerebral de cada paciente ocupará un espacio estándar. No obstante, para que el algoritmo que implementa SPM normalice correctamente, las imágenes deben ser razonablemente similares a la plantilla usada, tanto morfológicamente como en contraste (en la práctica, en el caso de imágenes cerebrales basta con que tengan la misma orientación y la misma profundidad de bits). La metodología usada por SPM es equivalente a la descrita en [Evans *et al.*, 1993].

SPM proporciona plantillas para diferentes tipos de imágenes, todas ellas adquiridas en el Instituto Neurológico de Montreal [Evans *et al.*, 1993]. La plantilla para PET es una media de 12 imágenes de controles obtenidas usando el trazador $H_2^{15}O$. Para la de SPECT se toma la media de 22 imágenes de mujeres sanas adquiridas mediante el trazador 99^mTc -HMPAO. Estas plantillas suelen usarse para normalizar imágenes PET y SPECT independientemente del trazador usado, si bien, lo ideal sería usar una plantilla específica para cada trazador. El problema es que las imágenes que se usen para construir una plantilla específica para un trazador deben estar normalizadas previamente por lo que en algún momento en el proceso es necesario usar las plantillas estándar que proporciona SPM.

La normalización espacial posibilita la comparación vóxel a vóxel de las imágenes y facilita la localización de las regiones funcionales. SPM no comprueba que la normalización sea correcta por lo que es recomendable que el usuario compruebe visualmente que la plantilla y la imagen normalizada tienen la misma forma y el mismo tamaño. Es importante destacar que la normalización espacial iguala el tamaño y la forma de las imágenes pero no elimina las diferencias debidas a las características metabólicas de cada individuo. Además, las diferencias debidas al ruido también permanecen por lo que es necesario aplicar una etapa

de filtrado posterior. Por otro lado, el hecho de que las mismas regiones realicen la misma función en diferentes sujetos es cuestionable [Donoghue, 2002; Cekic *et al.*, 2007].

6.2.3. Normalización en intensidad

El metabolismo cerebral medio varía significativamente entre diferentes pacientes e incluso en el mismo paciente en diferentes instantes. Por esta razón es necesario normalizar los niveles de intensidad de las imágenes funcionales antes de realizar análisis estadísticos sobre ellas. Esta normalización pueden llevarse a cabo de diferentes formas:

- **Escalado.** Consiste en multiplicar la intensidad de cada imagen por un factor de modo que se iguale la intensidad media de todas las imágenes. Este enfoque es el más simple y adecuado para estudios con imágenes procedentes de distintos pacientes.
- **Análisis de covarianza (ANCOVA).** Consiste en incluir la intensidad media de cada imagen en el modelo estadístico de forma que las diferencias en intensidad sean tenidas en cuenta durante el análisis. Este enfoque realiza un escalado aditivo que es mejor que el escalado proporcional cuando la varianza entre diferentes imágenes es baja como es el caso de los estudios con un solo paciente.

Estos dos procedimientos están implementados en SPM y no existe consenso sobre cual es mejor. Sin embargo, ninguno de ellos es apropiado cuando los efectos de interés afectan a áreas suficientemente grandes como para influir en el valor de intensidad medio de la imagen. En este caso es preferible usar otros enfoques como los que se describen a continuación:

- **Normalization basada en regiones.** Este método es similar al escalado proporcional. Escala las imágenes igualando las regiones no afectadas por la patología bajo estudio. Su principal inconveniente es la necesidad de identificar y delimitar las regiones no afectadas.
 - **Normalización con respecto al máximo.** Iguala la intensidad máxima de las imágenes y escala los vóxeles de cada imagen con respecto al máximo de esa imagen. Este método implica asumir que las regiones con mayor metabolismo no se ven afectadas por la patología bajo estudio.
-



Figura 6.2: Suavizado de imágenes. *Izquierda*: Imagen original. *Derecha*: Imagen suavizada

6.2.4. Filtrado espacial

El filtrado espacial es un proceso por el cual los vóxeles son promediados con sus vecinos produciendo un efecto de suavizado cuya intensidad depende de un parámetro llamado FWHM (Full Width at Half Maximum). Este parámetro se mide en unidades de espacio y normalmente se le da un valor igual a tres veces el tamaño del vóxel. La figura 6.2 ilustra el proceso de suavizado sobre una imagen bidimensional. El filtrado espacial mejora el ajuste de la teoría de campos Gaussianos [Worsley *et al.*, 1996] en la que se basan los análisis de SPM pero además tiene otros objetivos. Por un lado, incrementa la relación señal ruido dado que elimina ruido de la imagen. Por otro lado asegura que las diferencias entre pacientes se dan a tamaños mayores que el de un vóxel garantizando así que las diferencias sean significativas.

6.3. Análisis estadísticos

SPM permite realizar varios tests estadísticos: test t de Student, tests F y análisis de varianza (Anova) incluyendo variables y interacciones de unas varia-

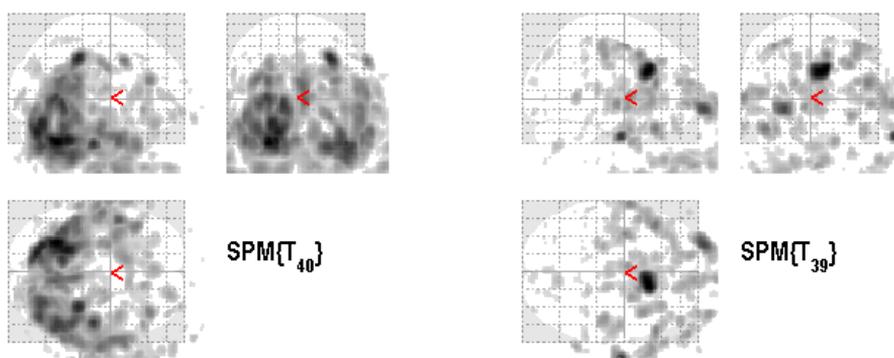


Figura 6.3: *Izquierda*: Mapa de activación de un estudio categórico que compara un conjunto de controles con un enfermo de AD. *Derecha*: Mapa de activación de un estudio categórico que compara un conjunto de controles con una persona sana. La hipótesis en ambos casos es que la imagen bajo estudio tiene menor activación que el conjunto de controles.

bles con otras. Estos tests se engloban en una formulación común conocida como el Modelo Lineal General (GLM) [Friston *et al.*, 1994]. Este modelo introduce dos conceptos básicos, la matriz de diseño y los contrastes. Es importante destacar que SPM, así como otros estudios estadísticos no busca efectos directamente en las imágenes funcionales sino que únicamente acepta o rechaza una hipótesis definida *a priori*. Por tanto, un estudio SPM debe tener una hipótesis de partida sobre las razones (o efectos) que producen el comportamiento observado. Esta hipótesis se conoce como hipótesis nula y se representa mediante la matriz de diseño.

Los estudios estadísticos que realiza SPM se dividen en dos grupos: estudios paramétricos o factoriales y estudios categóricos o sustractivos. Los estudios paramétricos analizan la relación entre las imágenes y un parámetro como la edad o el resultado de un test cognitivo. Esta relación debe ser lineal o, de lo contrario, es necesario llevar a cabo una transformación que la linealice. Los estudios categóricos se usan para resaltar las diferencias entre dos grupos de imágenes. Por ejemplo, la hipótesis de que la edad influye en el metabolismo cerebral puede ser aceptada o rechazada mediante un test paramétrico. Esta hipótesis daría lugar a una matriz de diseño con una fila para cada imagen incluida en el estudio y una columna para la edad. Generalmente la matriz

de diseño de un estudio paramétrico tiene una columna por cada efecto que se quiera estudiar. La matriz de diseño de los estudios categóricos incluye una columna adicional para representar la pertenencia de cada imagen a uno u otro grupo. Existe un tercer grupo que recoge aquellos estudios que combinan un estudio paramétrico y un estudio categórico. Por ejemplo supongamos que la hipótesis nula consiste en que ciertas regiones presentan menor metabolismo en el grupo patológico que en grupo de controles (efecto de interés) después de eliminar el efecto de la edad (efecto regulador). En este caso la matriz de diseño tendría tres columnas, dos para indicar la pertenencia a un grupo y otra para indicar la edad.

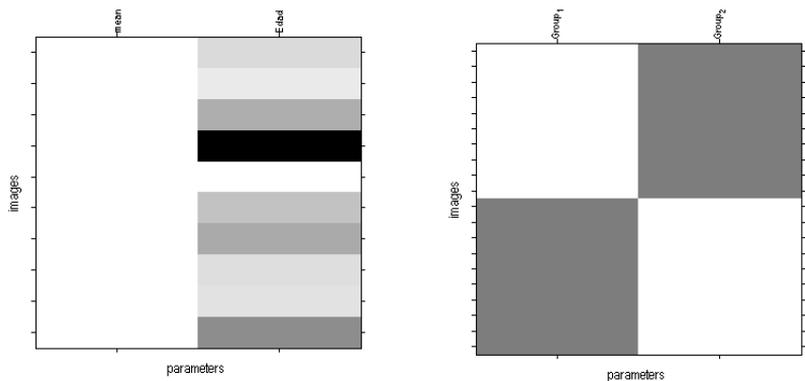
En la matriz de diseño todos los efectos están al mismo nivel, es decir, todos se representan mediante una columna, sin embargo, no todos los efectos son iguales. En el ejemplo anterior distinguimos entre efectos de interés y efectos reguladores. El vector de contraste permite realizar esta diferenciación. Un vector de contraste es un vector de longitud igual al número de efectos incluidos en la matriz de diseño. Cada efecto regulador se representa con un 0 en el vector de contraste. Los efectos paramétricos se representan con un 1 si la correlación es positiva y con un -1 si la correlación es negativa. En caso de efectos categóricos la suma de todos los elementos del vector de contraste debe ser 0.

La figura 6.4 muestra tres matrices de diseño correspondientes a tres estudios diferentes. La matriz de diseño representada en la figura 6.4(c) tendría un vector de contraste $c = (-1, 1, 0)$. De acuerdo con esta hipótesis, los pacientes de AD tendrían menor metabolismo que los controles, por eso su valor en el vector de contraste es negativo. Por el mismo motivo, el valor para la columna de controles es positivo (los controles tienen mayor metabolismo que los pacientes de AD). Los dos primeros valores se corresponden con efectos categóricos y deben sumar 0. El tercer valor corresponde a un efecto regulador y por definición debe ser 0.

6.4. Mapa de estadísticos paramétricos

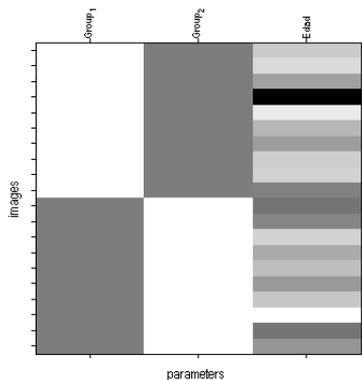
Después de definir la matriz de diseño y el vector de contrastes es posible realizar el test estadístico (un test t de Student o un test F) para cada vóxel de la imagen de forma independiente. El resultado es una nueva imagen conocida como mapa de estadísticos paramétricos y que contiene el resultado del test (valor p) en cada vóxel. Los valores p representan la probabilidad de que no haya efectos de interés significativos.

Las imágenes funcionales tienen una gran cantidad de vóxeles, lo que se traduce en un alto número de valores p , ya que éstos son analizados indepen-



(a) Análisis de la relación entre la actividad cerebral y la edad de los pacientes (estudio paramétrico).

(b) Diferencias entre la actividad cerebral de pacientes con AD y controles (estudio categórico).



(c) Diferencias entre la actividad cerebral de pacientes con AD y controles, considerando la influencia de la edad (estudio mixto).

Figura 6.4: Matrices de diseño de tres estudio en SPM.

dientemente. Por este motivo siempre aparecerán algunos valores p que exceden un umbral preestablecido, debido simplemente al azar. Más concretamente, si no hay efectos en las imágenes, el 1% serán menores de 0.01, el 0.1% serán menores que 0.001, etc. Este problema se conoce como el problema de las múlti-

ples comparaciones y es el culpable de los falsos positivos. Habitualmente se resuelve mediante la corrección de Bonferroni [Chatfield & Collins, 1980]. Con ella, el valor p a partir del cual se acepta la hipótesis nula se calcula como $\alpha/\text{numero_de_tests}$ donde α es la tasa de falsos positivos que estamos dispuestos a admitir. Sin embargo la corrección de Bonferroni asume que los tests son independientes unos de otros y esto no ocurre en el caso de imágenes funcionales. El valor de un vóxel en una imagen funcional es, en la mayoría de los casos, muy similar al de los vóxeles vecinos. Además hay otras razones por las que siempre aparecen algunos grados de correlación espacial en las imágenes:

- La forma de capturar y reconstruir la imagen.
- Las señales psicológicas afectan a áreas extensas.
- El preprocesado espacial que se aplica a las imágenes antes del análisis (realineamiento, normalización espacial, suavizado, etc.).

SPM usa la Teoría de Campos Gaussianos para resolver el problema de la correlación de los vóxeles de un modo más adecuado que la corrección de Bonferroni.

Aunque generalmente se acepta que un umbral $p < 0,05$ (con el que se obtiene un falso positivo cada 20 observaciones) es significativo, la interpretación de los valores p no es trivial [Goodman, 1999]. Actualmente es una práctica común realizar test de permutación y remuestreo para cada conjunto de imágenes con el fin de validar la significancia de los umbrales para los valores p [Ashburner, 2000; Arndt *et al.*, 1996]. Estos procedimientos implican repetir varias veces el test estadístico reasignando imágenes aleatoriamente a cada grupo (bootstrap) o eliminando algunas de imágenes (jackknife). Además, para que un valor p sea aceptado como significativo debe llevar a una hipótesis plausible, basada en el conocimiento sobre los mecanismos biológicos implicados en los procesos bajo estudio. De esta forma se pueden explicar los resultados de forma consistente.

6.5. Limitaciones de SPM

SPM asume que los datos se adecuan a GLM, lo que implica que los efectos de interés sean aditivos y la distribución de los residuos sea una Gaussiana de media cero e igual varianza entre grupos. Esta distribución se puede estimar mediante un diagrama Q-Q. Si la distribución es complemente Gaussiana, el diagrama mostrará una línea recta. En la práctica, aunque esta afirmación no se

satisface completamente, el método es lo suficientemente robusto para permitir la presencia de pequeñas variaciones.

Un problema mayor aparece cuando se estiman simultáneamente los efectos de factores que no son linealmente independientes. La dependencia produce resultados inestables: pequeños cambios en los datos producen resultados totalmente diferentes.

Por otro lado, el sistema de coordenadas usado para localizar las observaciones es a veces confuso. SPM usa un sistema de coordenadas basado en el atlas de Talairach pero los valores proporcionados por SPM no son equivalentes a los que proporciona Talairach. Esto es debido a que el atlas de Talairach está referido al cerebro de un solo paciente mientras que el sistema de coordenadas de SPM se refiere a un patrón construido como la media de varios pacientes.

Estudios más extensos sobre las limitaciones de SPM se pueden encontrar en [Ashburner, 2000; Missimer *et al.*, 1999; Davatzikos *et al.*, 2001; Lahorte *et al.*, 2000; Stamatakis *et al.*, 1999; Wang *et al.*, 1999].

6.6. Sistemas CAD basados en SPM

Aunque no es su principal propósito, SPM se puede usar para la construcción de sistemas CAD. Más concretamente, es posible usarlo para tareas de clasificación y como método de extracción de características [Segovia *et al.*, 2010b].

Así, por ejemplo, para diagnosticar AD a partir de la imagen cerebral de un paciente del cual sospechamos que padece la enfermedad, podemos usar un test t de Student junto con un conjunto de imágenes de personas sanas (controles). Una vez que las imágenes han sido preprocesadas, realizamos dicho test sobre las dos poblaciones compuestas, una de ellas, por los controles de la base de datos, y la otra, por la imagen bajo estudio. Tras obtener los mapas de estadísticos resultantes del análisis, podemos clasificar directamente la imagen en base a esa información. Si el mapa contiene un número suficientemente alto (n) de valores p que superan un umbral (t) es indicativo de que las poblaciones son muy distintas y por tanto pertenecen a clases distintas, es decir, la imagen bajo estudio será considerada como AD puesto que es *muy distinta* a los controles. Los resultados obtenidos mediante este enfoque son pobres debido fundamentalmente a que no hemos incluido en el modelo la información relativa al patrón de AD (usamos una base de datos con controles solamente). Si bien, también es cierto que esa información se incluye en parte en los parámetros del modelo, n y t cuyo cálculo no está exento de dificultad.

Por otro lado, es posible usar los mapas de estadísticos obtenidos con SPM

como método de extracción de características. Para ello debemos realizar el test estadístico (test t de Student por ejemplo) sobre las dos poblaciones formadas por controles e imágenes patológicas respectivamente (en este caso si que necesitamos una base de datos que contenga tanto controles como imágenes de pacientes que sufren la enfermedad que queremos diagnosticar). Del mapa de estadísticos resultante debemos seleccionar aquellos valores p que superen un cierto umbral y éstos serán usados como mascara para seleccionar ciertos vóxeles de las imágenes funcionales. Posteriormente, con la ayuda de un clasificador estadístico como los descritos en el capítulo 5 podemos construir un sistema CAD completo. Este modelo supondría un enfoque mixto, univariado en la extracción de características y multivariado en la clasificación posterior.

Parte III

Aportaciones de la Tesis
Doctoral

CAPÍTULO

7

ANÁLISIS DE IF BASADO EN EL MODELO DE MEZCLA DE GAUSSIANAS

Los modelos basados en agrupamiento o *clustering* [Jain *et al.*, 1999; Xu & Wunsch, 2005] han sido ampliamente usados en campos como el de reconocimiento de patrones [Anderberg, 1973], detección de voz [Górriz *et al.*, 2006a,b] o segmentación de imágenes [Jain & Flynn, 1996]. La idea básica consiste en agrupar datos que son similares en cierto sentido formando subconjuntos o *clusters*. Por ejemplo, en el caso de imágenes tridimensionales en color, un *cluster* estaría formado por varios vóxeles adyacentes del mismo color. En este capítulo veremos un nuevo enfoque para la extracción de características de imágenes funcionales basado en un modelo de *clustering*, el Modelo de Mezcla de Gaussianas.

7.1. Introducción

El uso de métodos de aprendizaje estadístico para la construcción de sistemas CAD conlleva la necesidad usar algoritmos que reduzcan la dimensionalidad de los datos de entrada a fin de evitar el problema del pequeño tamaño muestral. Por otro lado, para el problema que se aborda en esta tesis, la alta variabilidad de las clases supone una dificultad añadida que hace que técnicas clásicas de reducción de la dimensionalidad como PCA o ICA seleccionen como características información no relevante para la clasificación y que redundará en tasas de acierto mejorables.

En este sentido, los métodos de *clustering* o parcelación proporcionan una nueva vía para conseguir algoritmos de extracción de características más eficaces. Estos métodos se usan con frecuencia en problemas de segmentación de datos y compresión [Jain *et al.*, 1999; Xu & Wunsch, 2005]. La idea básica consiste en agrupar puntos de datos que son similares en algún sentido formando subconjuntos o parcelas. Por ejemplo, en el caso de imágenes de dos dimensiones, las parcelas pueden ser píxeles continuos del mismo color (si las imágenes son en color) o con la misma intensidad de gris (si se trata de imágenes en escala de grises). Recientemente se han usado métodos de parcelación o *clustering* basados en mezcla de Gaussianas para cuantificar la distribución espacial del color en imágenes [Goldberger *et al.*, 2006]. Además las técnicas de *clustering* se han aplicado satisfactoriamente en varios campos como reconocimiento de patrones [Anderberg, 1973], detección de voz [Górriz *et al.*, 2006a,b] o segmentación de imágenes [Jain & Flynn, 1996]. En estudios de imágenes funcionales, diferentes modelos basados en parcelación o *clustering* han sido empleados en análisis fMRI para agrupar coordenadas relevantes en el espacio Talarachich [Newman *et al.*, 2008]. Para esta tarea, se empleó en primer lugar la Estimación de Probabilidad de Activación (ALE, del inglés Activation Likelihood Estimation) para reducir la máximos de activación que tienen otros máximos en su vecindario. Después cada coordenada, \mathbf{x}_i , y su pertenencia a un *cluster*, \mathbf{z}_i , son ajustadas como una mezcla finita de distribuciones de probabilidad. El principal inconveniente de este método, derivado del uso de ALE, es que los patrones de hipoperfusión o hipometabolismo no están incluidos en el modelo. La aplicación de un modelo basado en análisis de imágenes debe tener en cuenta no sólo las coordenadas de los máximos de activación sino también los valores de intensidad de cada vóxel.

En este capítulo se presenta un enfoque diferente de parcelación que usa el modelo de mezcla de Gaussianas para estimar un perfil de intensidad de una imagen funcional que permita reducir la dimensión de los vectores de características correspondientes a estas imágenes con el fin de construir sistemas

CAD más eficaces. Este enfoque modela el perfil de intensidades de una imagen funcional como una suma de Gaussianas que satisface un criterio de máxima probabilidad. Para este propósito se usa el algoritmo de maximización de la esperanza debido a su simplicidad y a su robustez [Newman *et al.*, 2008], si bien, también pueden usarse otros métodos de maximización. La idea es agrupar los vóxeles similares de una imagen funcional para que en el vector de características correspondiente solamente aparezca un elemento por cada grupo, en lugar de uno por cada vóxel (como se hace en VAF) reduciendo de esta forma el tamaño de los vectores de características. Para extraer los grupos o *clusters* subyacentes de una imagen funcional, usamos en primer lugar una imagen construida como la media varios controles. Mediante la aplicación del modelo de mezcla de Gaussianas, cada uno de estos grupos es modelado por una Gaussiana con una determinada media (posición del grupo), covarianza (forma del grupo en el espacio) y peso (relativo a la intensidad de los vóxeles que forman el grupo). Los vectores de características de cada imagen se construyen como la activación que tiene dicha imagen en las regiones que define cada Gaussiana. Por tanto el número de características será igual al número de Gaussianas usadas para modelar una imagen funcional. La reducción del espacio de características de esta forma permite conservar tanto la especificidad regional (por ejemplo, regiones de hipoperfusión o hipometabolismo) como los cambios globales en el mapa de activación del cerebro.

Una alternativa a la propuesta anterior consiste extraer los *clusters* a partir de una imagen con las diferencias entre las imágenes patológicas (imágenes de AD en nuestros experimentos) y los controles. De esta forma, el modelado mediante mezcla de Gaussianas cobra una nueva dimensión puesto que las Gaussianas modelan en este caso en las regiones de interés (ROIs, del inglés Region Of Interest) para el diagnóstico de la patología bajo estudio. Así se consigue que la extracción de características se centre sólo en las regiones que son importantes para discernir si un paciente padece la enfermedad o no, lo que mejorará las tasas de acierto en la fase de clasificación dando lugar a sistemas CAD más precisos.

La figura 7.1 muestra una representación en tres dimensiones de las regiones de interés para el diagnóstico de la AD. Las elipses representan Gaussianas y el color de las mismas es relativo a su peso, es decir, a la intensidad de los vóxeles que agrupa, relativa a la importancia de la ROI que representa.

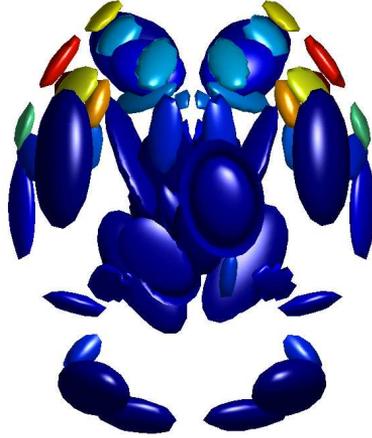


Figura 7.1: Vista superior la representación en tres dimensiones de las regiones de interés para el diagnóstico de AD modeladas como una mezcla de Gaussianas.

7.2. Definición del modelo

Los Modelos de Mezcla de Gaussianas (GMMs, del inglés Gaussian Mixture Models) son quizás los métodos estadísticos más maduros para realizar *clustering* o agrupamiento (véase por ejemplo [Fukunaga, 1990]) y se usan frecuentemente para estimación de densidad [McLachlan & Peel, 2000; Aladjem, 2005].

La idea básica de GMM para estimación de densidad es que los datos que queremos modelar, \mathbf{x}_i , $i = 1 \dots N$, son observaciones de una distribución de probabilidad $p(\mathbf{x})$, que se modela mediante la suma de k Gaussianas.

$$p(\mathbf{x}) = \sum_{n=1}^k w_n f_n(\mathbf{x}|\theta_n) \quad (7.1)$$

donde $f_n(\mathbf{x}|\theta_n)$ es la densidad de la Gaussiana n con vector de parámetros θ_n y los w_n son los pesos de las Gaussianas con $\sum_{n=1}^k w_n = 1$. Las distribuciones normales $f_n(\mathbf{x}|\theta_n)$ en d dimensiones vienen dadas por

$$f_n(\mathbf{x}|\theta_n \in \{\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n\}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_n|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_n)^T \boldsymbol{\Sigma}_n^{-1}(\mathbf{x}-\boldsymbol{\mu}_n)} \quad (7.2)$$

con valores esperados $\boldsymbol{\mu}_n$ y matrices de covarianza $\boldsymbol{\Sigma}_n$. Las características geométricas de las Gaussianas se pueden ajustar mediante las matrices de covarian-

za Σ_n usando la descomposición en valores propios [Banfield & Raftery, 1993]. Para el caso que nos ocupa, se asume que la forma, volumen y orientación de las Gaussianas es variable dado que las diferentes regiones de interés que se localizan en el cerebro no tienen la misma forma y tamaño.

7.3. Estimación por máxima verosimilitud

La estimación por máxima verosimilitud (MLE, del inglés Maximum Likelihood Estimation) proporciona una forma de calcular los parámetros del modelo de mezcla de Gaussianas (w_n , μ_n y Σ_n). Para ello un algoritmo MLE maximiza la siguiente expresión:

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{i=1}^N p(\mathbf{x}_i|\theta) \quad (7.3)$$

esto es:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta|\mathbf{x}) \quad (7.4)$$

donde $\theta = \{\theta_n\}$, para $n = 1, \dots, k$ y $\mathbf{x} = \{\mathbf{x}_i\}$, para $i = 1, \dots, N$ en un modelo con k Gaussianas, lo cual se corresponde con la probabilidad de observar los datos \mathbf{x}_i , asumiendo que las variables son aleatorias e idénticamente distribuidas [Newman *et al.*, 2008; McLachlan & Peel, 2000].

Si los datos se agrupan en un histograma con B barras en las posiciones \mathbf{x}_j , $j = 1 \dots B$, y con alturas h_j , la estimación de máxima verosimilitud puede usarse de distinta forma [McLachlan & Jones, 1988]. En este caso, el nivel de gris de cada observación es tenido en cuenta como parte de parámetro h_j y el número total de observaciones viene dado por:

$$N = \sum_{j=1}^B h_j \quad (7.5)$$

Partiendo de la ecuación 7.3 y usando la ecuación 7.1, la probabilidad queda como:

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{j=1}^B [p(\mathbf{x}_j)]^{h_j} = \prod_{j=1}^B \left[\sum_{n=1}^k w_n f_n(\mathbf{x}_j|\theta_n) \right]^{h_j} \quad (7.6)$$

dado que hay h_j observaciones en \mathbf{x}_j . Esta probabilidad es esencialmente diferente de enfoques previos [Newman *et al.*, 2008; Fraley & Raftery, 1998] dado que aquí, las alturas h_j no son parámetros del modelo que deban ser determinados y no están relacionados con la pertenencia a una determinada Gaussiana.

La maximización se consigue habitualmente maximizando el logaritmo de la probabilidad en vez de la probabilidad misma, ya que los dos máximos se dan para los mismos valores de los parámetros y el máximo del logaritmo de la probabilidad es, por lo general, más fácil de determinar. El logaritmo de la probabilidad se expresa como:

$$\ln \mathcal{L}(\theta|\mathbf{x}) = \sum_{j=1}^B h_j \ln \left(\sum_{n=1}^k w_n f_n(\mathbf{x}_j|\theta_n) \right) \quad (7.7)$$

donde las alturas del histograma, h_j , son usadas como pesos. Maximizar el logaritmo de la probabilidad bajo la restricción $\sum_{n=1}^k w_n = 1$ es equivalente a encontrar el máximo de la siguiente expresión:

$$\mathcal{J} = \sum_{j=1}^B h_j \ln \left(\sum_{n=1}^k w_n f_n(\mathbf{x}_j|\theta_n) \right) - \lambda \left(\sum_{n=1}^k w_n - 1 \right), \quad (7.8)$$

donde el primer término es el logaritmo de la probabilidad y λ es un multiplicador de Lagrange que asegura una correcta normalización.

Por otro lado, si en cada experimento se producen una serie de datos observables, \mathbf{x} , y otros no observables, \mathbf{z} , debemos maximizar la probabilidad del conjunto completo de datos, $\mathbf{y} = (\mathbf{x}, \mathbf{z})$, y buscar $\hat{\theta}$ tal que:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta, \mathbf{z}|\mathbf{x}) \quad (7.9)$$

En este caso, el logaritmo de la probabilidad del conjunto completo de datos se expresa como [Newman *et al.*, 2008]:

$$\ell(\theta, \mathbf{z}|\mathbf{x}) = \sum_{i=j}^B \sum_{n=1}^k h_j z_{jn} \log(w_n f_n(\mathbf{x}_j, |\theta_n)) \quad (7.10)$$

donde los datos no observables $\mathbf{z}_i = (z_{i1}, \dots, z_{ik})$ representan la asociación del vector \mathbf{x}_i a una de las k Gaussianas. La suposición más relevante en la ecuación 7.10 es que la densidad de una observación \mathbf{x}_i dada \mathbf{z}_i viene dada por $\prod_{n=1}^k f_n(\mathbf{x}_i|\theta_k)^{z_{ik}}$ y cada \mathbf{z}_i es independiente e idénticamente distribuida de

acuerdo a una distribución multinomial de una repetición y k categorías con probabilidades w_1, \dots, w_k [Fraley & Raftery, 1998].

La estimación por máxima verosimilitud en presencia de datos no observables puede calcularse mediante un algoritmo de maximización de la esperanza cuyo funcionamiento se describe en la siguiente sección.

7.4. Algoritmo de maximización de la esperanza

El algoritmo de maximización de la esperanza (EM, del inglés Expectation-Maximization) [McLachlan & Jones, 1988] se utiliza con frecuencia dado que se trata un esquema recursivo con interesantes propiedades de convergencia monótona [Dempster *et al.*, 1977; Moon, 1996]. Supone un enfoque general para maximizar la probabilidad en presencia de datos no observables (véase ecuación 7.9) en un esquema MLE. Partiendo de una suposición inicial, el algoritmo EM procede alternativamente estimando los datos no observables, \mathbf{z} , y los parámetros, θ . Concretamente, en el paso E el algoritmo calcula el logaritmo de la probabilidad del conjunto completo de datos con respecto a \mathbf{z} dado \mathbf{x} y la actual estimación de θ . En el paso M, el valor del logaritmo de la probabilidad que se calculó en el paso E se maximiza en términos de θ , manteniendo \mathbf{z} fijo.

El valor de z_{in} en el máximo de la ecuación 7.10 es la probabilidad estimada de que \mathbf{x}_i esté asociado a la Gaussiana n y la clasificación de máxima probabilidad de \mathbf{x}_i es la Gaussiana $n = \arg \max_n \mathbf{x}_n z_{in}$ para $n = 1, \dots, k$.

En la misma línea que lo que se muestra por ejemplo en [Fukunaga, 1990] y usando la nueva probabilidad logarítmica que se muestra en la ecuación 7.7 podemos definir las ecuaciones para actualizar los parámetros w_n , $\boldsymbol{\mu}_n$ y $\boldsymbol{\Sigma}_n$ (donde las relaciones son sólo modificadas por los pesos h_j):

$$w_n = \frac{1}{N} \sum_{j=1}^B h_j q_n(\mathbf{x}_j) \quad (7.11)$$

$$\boldsymbol{\mu}_n = \frac{1}{w_n N} \sum_{j=1}^B h_j q_n(\mathbf{x}_j) \mathbf{x}_j, \quad (7.12)$$

$$\boldsymbol{\Sigma}_n = \frac{1}{w_n N} \sum_{j=1}^B h_j q_n(\mathbf{x}_j) (\mathbf{x}_j - \boldsymbol{\mu}_n)(\mathbf{x}_j - \boldsymbol{\mu}_n)^T \quad (7.13)$$

donde $q_n(\mathbf{x})$ es la probabilidad posterior y se define como:

$$q_n(\mathbf{x}) = \frac{w_n f_n(\mathbf{x})}{p(\mathbf{x})} \quad (7.14)$$

El algoritmo EM parte de unos valores iniciales para w_n , $\boldsymbol{\mu}_n$ y $\boldsymbol{\Sigma}_n$ y va aplicando las ecuaciones 7.11, 7.12 y 7.13 recursivamente hasta que converge, es decir, hasta que los cambios en el valor de la probabilidad logarítmica son menores que un umbral dado.

7.4.1. Algunos comentarios sobre el algoritmo EM

En general el algoritmo EM converge monótonamente [Dempster *et al.*, 1977], sin embargo se sabe que el procedimiento puede converger hacia un máximo local. En este sentido, la elección de los valores iniciales para los parámetros w_n , $\boldsymbol{\mu}_n$ y $\boldsymbol{\Sigma}_n$ juega un importante papel. En [Fraley & Raftery, 1998] se propone un método basado en la aglomeración jerárquica para alcanzar una agrupación razonable en ausencia de información inherente a los datos sobre una posible agrupación. Por razones computacionales, es preferible adaptar el procedimiento de inicialización a los datos a los que se va a aplicar, como se muestra en el siguiente apartado. Otra cuestión interesante acerca de esta técnica es la selección del modelo, esto es, determinar el número de grupos inherentes en los datos. Se han propuestos varios métodos para resolver los problemas de selección del modelo. Uno de ellos es el criterio de información bayesiano (BIC, del inglés Bayesian Information Criterion) [Schwarz, 1978], sin embargo, el más usado consiste en aplicar varios modelos con diferente número de componentes y elegir el mejor de acuerdo a algún criterio de selección (requerimientos computacionales, reducción de dimensionalidad, tamaño de los vectores de características en comparación con el número de muestras, precisión en la definición de las regiones de interés, etc) [Górriz *et al.*, 2006a; Goldberger *et al.*, 2006].

7.4.2. Inicialización

La inicialización del algoritmo EM es muy importante para evitar que se caiga en un máximo local. Al aplicarlo a imágenes funcionales, la caída en un máximo local es especialmente importante ya que la configuración de las Gaussianas podría evolucionar hacia configuraciones cualitativamente diferentes correspondientes a diferentes máximos locales aun cuando las imágenes fuesen muy similares unas a otras.

Sin embargo, en nuestro caso, el algoritmo EM solamente se aplica a una imagen inicial (véase sección 7.5), por lo que no es esencial encontrar la probabilidad máxima global. Para nuestro propósito es recomendable colocar las k Gaussianas centradas en una malla regular con $n_x \times n_y \times n_z$ cuadros. Con el fin de no poner los centros de las Gaussianas en un máximo local es preferible colocarlos en regiones de baja activación para que a partir de ellas evolucionen libremente. Por lo tanto, nosotros utilizamos el cubo cuyo lado es $\frac{1}{3}$ de la longitud del lado de la imagen, donde la intensidad es bastante baja, y distribuimos los centros de Gaussianas regularmente por esta región. La distancia entre los puntos de la cuadrícula es de $l_i = L_i/(3n_i)$, con $i = x, y, z$ y donde L_i es la longitud de la imagen en la dirección i . Los pesos se inicializan con $w_n = 1/k$, y las matrices de covarianza son elegidas como matrices diagonales con elementos $3l_x$, $3l_y$ y $3l_z$. Este esquema proporciona una configuración de inicialización regular y simétrica, donde las Gaussianas inicialmente cubren todo el cubo central de la imagen cerebral y están considerablemente superpuestas.

7.4.3. Selección del modelo

El criterio de selección del modelo es elegido para ser dependiente de los datos, dado que con un número k suficiente de Gaussianas, todas las áreas de activación relevantes serán incluidas en él. Nosotros usamos un valor de $k = 64$ Gaussianas inicializadas en una malla de $4 \times 4 \times 4$. Este número permite una configuración cúbica con igual número de Gaussianas centradas en cada dirección que resultó ser la mejor opción para el esquema de inicialización: las Gaussianas resultantes son pequeñas y la imagen puede ser reconstruida muy bien (véase figura 7.2) pero aún la dimensionalidad de los vectores de características, directamente relacionada con el número de Gaussianas, se mantiene en un valor aceptable. Otros modelos han sido aplicados pero no cumplen algunos criterios de selección de los mencionados anteriormente, así, usando por ejemplo $3^3 = 27$ Gaussianas, las ROIs son más grandes, lo que lleva a resultados de clasificación más pobres y a una peor reconstrucción de la imagen original a partir de las Gaussianas. Con $10^3 = 1000$ Gaussianas la dimensionalidad resultante, es decir, el tamaño de los vectores de características es muy elevado (mayor que el número de muestras en la mayoría de los experimentos) con lo que aparece el problema del tamaño muestral [Duin, 2000], lo que igualmente redundaría en pobres resultados de clasificación.

Una forma de medir la adecuación del modelo para representar una imagen es mediante el error de reconstrucción. La figura 7.3 muestra un estudio de la capacidad de reconstrucción de una imagen funcional a partir de su representa-

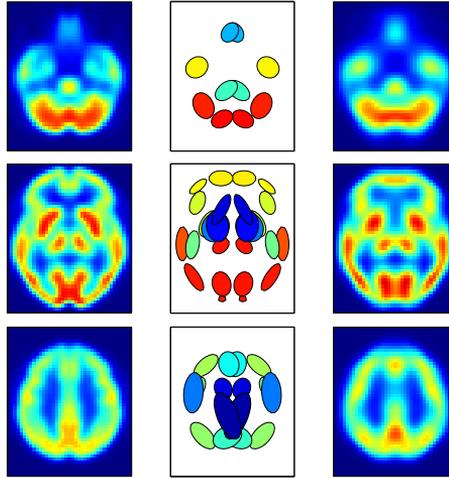


Figura 7.2: Columna izquierda: diferentes cortes de una imagen SPECT construida como la media de varios controles. Columna central: Representación mediante Gaussianas (el color de cada elipse representa la intensidad de la Gaussiana). Columna derecha: Imagen reconstruida a partir de las Gaussianas

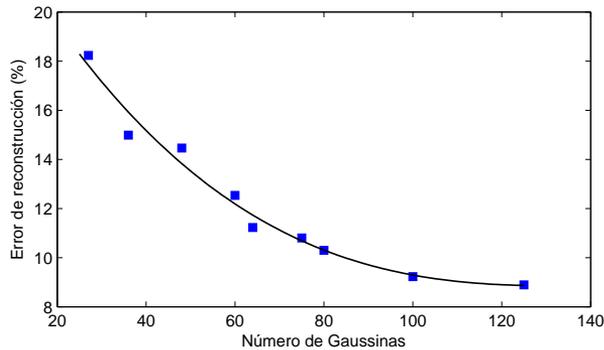


Figura 7.3: Error de reconstrucción de una imagen SPECT a partir de su representación mediante Gaussianas en función del número de Gaussianas usadas en el modelo

ción mediante Gaussianas. En la figura se representa el error de reconstrucción, \mathbf{E}_{rec} , en función del número de Gaussianas usadas en el modelo.

$$\mathbf{E}_{rec} = \frac{\sqrt{\frac{\sum_{i=1}^n (\mathbf{I}_i - \mathbf{I}_i^{rec})^2}{n}}}{I_{max}} \quad (7.15)$$

donde \mathbf{I}_i e \mathbf{I}_i^{rec} es el i -ésimo vóxel de la imagen original y la imagen reconstruida respectivamente, n es número de vóxeles e I_{max} es la intensidad máxima. Esta figura muestra como el error de reconstrucción tiende a estabilizarse en torno al 9% cuando el número de Gaussianas es alto por lo que no tiene mucho sentido usar más de 100 Gaussianas. Por otro lado, un número de Gaussianas muy pequeño da lugar a un error de reconstrucción demasiado alto.

7.5. Extracción de características

La extracción de características, en este contexto, es el paso a partir del cual una imagen funcional del cerebro (con algo más de 500000 vóxeles una vez normalizada) se transforma en un vector de dimensiones reducidas (idealmente de menos de 100 elementos) para poder usarse posteriormente junto a técnicas de aprendizaje automático sin que se produzca el problema del pequeño tamaño muestral.

De acuerdo con el modelo propuesto en este capítulo, el proceso de extracción de características puede dividirse en 3 etapas:

1. Construcción de la imagen modelo, I_M , que posteriormente representaremos mediante una mezcla de Gaussianas.
2. Aplicación del algoritmo de maximización de la esperanza para obtener la mezcla de Gaussianas subyacente en I_M .
3. Composición del vector de características de una imagen funcional a partir de la configuración de Gaussianas obtenida en el paso anterior.

Una primera aproximación de este método consiste en construir la imagen I_M como la media de varias imágenes funcionales. En este caso, hemos usado solamente controles para construir la media debido a que el patrón de hipoperfusión/hipometabolismo de las imágenes patológicas es muy variable. Sin embargo, dado que el objetivo último es la clasificación y por tanto la diferenciación entre controles e imágenes patológicas resulta más interesante definir la

imagen I_M de otra manera. Una segunda aproximación que redundará en sistemas CAD más eficaces consiste en construir I_M como la diferencia entre los controles y las imágenes patológicas:

$$I_M = I_C - I_P \quad (7.16)$$

donde I_C e I_P es la media de los controles e imágenes patológicas respectivamente. De esta forma conseguimos que el algoritmo se centre sólo en las regiones que son importantes para la clasificación.

La siguiente etapa consiste en aplicar el algoritmo EM para obtener la representación mediante Gaussianas de I_M cuyo perfil de intensidades se asume que es una suma de Gaussianas. Además, asociado a cada Gaussiana se calcula un peso que indica la importancia de dicha Gaussiana en la mezcla y que representa la intensidad de los vóxeles que agrupa. Estos pesos suponen una estimación la importancia de la ROI a la que representa cada Gaussiana.

En la figura 7.4 se muestra la imagen I_M construida como la diferencia entre controles e imágenes patológicas a partir de las imágenes SPECT de la base de datos del hospital Virgen de las Nieves (sec. 3.1). El color de las Gaussianas representa su peso en la mezcla. Nótese como las regiones que aparecen en la imagen coinciden con las que se nombran en la literatura como regiones representativas de la enfermedad de Alzheimer [de Leon *et al.*, 1983].

Finalmente, la configuración de Gaussianas obtenidas se aplica de forma similar a una mascara para construir el vector de características, $v_{\mathbf{x}}$, de cada imagen funcional.

$$v_{\mathbf{x}} = (c_1, \dots, c_k) \quad (7.17)$$

donde c_n representa el nivel de activación de la Gaussiana n para la imagen \mathbf{x} y viene dado por:

$$c_n = h_n \sum_{i=1}^V I(\mathbf{x}_i) f_n(\mathbf{x}_i) \quad (7.18)$$

donde $h_n = \frac{w_n}{\sqrt{(2\pi)^3 |\Sigma_n|}}$ y f_n son la altura y la densidad de la Gaussiana n respectivamente. \mathbf{x}_i son las coordenadas en el espacio del i -ésimo vóxel (con $i = 1, \dots, V$) de la imagen \mathbf{x} y $I(\mathbf{x}_i)$ es la intensidad del vóxel \mathbf{x}_i .

Por tanto, el número de características para cada imagen es igual al número de Gaussianas que usamos en el modelo. Si bien, es posible reducir más aún la dimensionalidad del problema truncando los vectores $v_{\mathbf{x}}$. Para ello sus

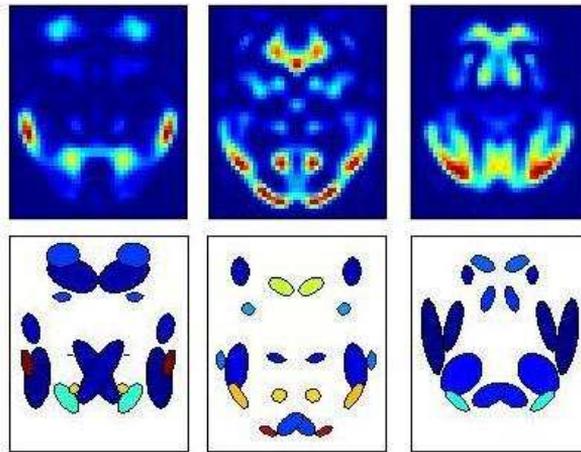


Figura 7.4: Regiones de interés para imágenes SPECT. Fila superior: Tres cortes de la imagen de diferencias I_M . Fila inferior: Representación mediante Gaussianas de la misma imagen (el color representa el peso de la Gaussiana) [Segovia *et al.*, 2010a]

componentes (los c_n) deben haber sido ordenados de acuerdo al peso w_n de la Gaussiana a la que representan, de forma que los que ocupan las primeras posiciones se corresponden con las Gaussianas de mayor peso.

CAPÍTULO

8

ANÁLISIS DE IF BASADO EN MÍNIMOS CUADRADOS PARCIALES

Por Mínimos Cuadrados Parciales (PLS, del inglés Partial Least Squares) se entiende un amplio grupo de métodos para el modelado de relaciones entre conjuntos de variables. PLS comprende tareas de regresión y clasificación así como técnicas de reducción de la dimensionalidad y herramientas de modelado. La idea subyacente de todos los métodos basados en PLS es que los datos observados son generados por un sistema que está dirigido por un pequeño número de variables latentes (que no es posible medir directamente puesto que no son observables) [Rosipal & Krämer, 2006]. PLS ha atraído la atención de la comunidad científica gracias a su exitosa aplicación en numerosas áreas como bioinformática, investigación de alimentos, medicina, farmacología, etc. [Nguyen & Rocke, 2002; Worsley, 1997; Nilsson *et al.*, 1997; Hulland, 1999].

8.1. Introducción

El método PLS [Varmuza & Filzmoser, 2009] original fue desarrollado en 1975 por el estadista Herman Wold para el tratamiento de cadenas de matrices y aplicaciones económicas. Posteriormente, su hijo Svante Wold y otros extendieron la idea de PLS a otros campos. No obstante, PLS siguió siendo un desconocido para gran parte de la comunidad científica durante mucho tiempo, y la mayoría de los estadistas lo veían más como un algoritmo que como un modelo estadístico riguroso [Helland, 2001]. Sin embargo, en los últimos años, el interés en PLS ha crecido hasta cotas similares al interés que despiertan modelos clásicos de regresión como Principal Component Regression (PCR) [Massy, 1965] o Ridge Regression (RR) [Hoerl & Kennard, 1970]. La efectividad de PLS ha sido estudiada teóricamente en términos de su varianza [Phatak *et al.*, 2002] y sus propiedades de contracción [Goutis, 1996]. De la misma forma, el rendimiento de PLS ha sido estudiado en varios trabajos [Frank & Friedman, 1993; Almoy, 1996].

En su forma general, PLS crea vectores conocidos como *scores* (también llamados vectores latentes o componentes) maximizando la covarianza entre diferentes conjuntos de variables. En este trabajo se hace uso de la variante más general de PLS que utiliza dos bloques de variables, si bien existen extensiones de este modelo general para manejar un mayor número de conjuntos [Wold, 1982; Wangen & Kowalsk, 1989; Smilde *et al.*, 2004]. En cierto sentido, PLS es similar a técnicas clásicas como Análisis de Componentes Principales (PCA) o Análisis de Correlación Canónica (CCA) donde también se usa el concepto de vectores latentes [Mardia *et al.*, 1997]. Esencialmente, los esquemas de PLS y PCA para regresión son los mismos. Primero se transforman las variables observadas, \mathbf{X} , en un conjunto intermedio de variables latentes (componentes) y esas nuevas variables son usadas para la regresión con una variable dependiente, \mathbf{Y} . Ambos, PLS y PCA, hacen uso de los vectores latentes para la regresión, si bien, estos vectores en el caso de PCA se calculan a partir de \mathbf{X} únicamente mientras que en el caso de PLS también se tiene en cuenta el conjunto \mathbf{Y} . El criterio para el cálculo de los vectores latentes más usado en PLS es el de máxima covarianza entre *scores* e \mathbf{Y} (o entre *scores* en \mathbf{X} y *scores* en \mathbf{Y}). Este criterio combina alta varianza de \mathbf{X} y alta correlación, lo cual hace de PLS un término medio entre PCA (máxima varianza) y el método clásico de mínimos cuadrados (máxima correlación) [Varmuza & Filzmoser, 2009].

Los principales conceptos de PLS son:

- Puede ser fácilmente extendido como un método de regresión lineal muy
-

potente y que permite trabajar con un gran número de variables.

- El modelo resultante predice una o más propiedades, \mathbf{Y} , a partir de las variables dependientes originales, \mathbf{X} .
- Durante el desarrollo, un número relativamente pequeño de las componentes PLS calculadas son usadas internamente para la regresión.
- El número de componentes PLS determina la complejidad del modelo y se puede optimizar para elevar la capacidad de predicción del algoritmo.

PLS puede ser aplicado también a problemas de clasificación usando una matriz que codifique la pertenencia de un vector a una determinada clase. Un clasificador basado en PLS es, en cierta forma, similar a un clasificador lineal de Fisher [Barker & Rayens, 2003]. PLS puede ser usado además como una herramienta de reducción de la dimensionalidad, como lo es el Análisis de Componentes Principales (PCA). Para ello, se usan los vectores latentes como vectores de características. Finalmente, cabe destacar que la potente maquinaria de aprendizaje basada en *kernels* (véase sección 5.2) también se puede aplicar a PLS [Scholkopf & Smola, 2002].

8.2. Desarrollo matemático

PLS se describe a menudo como un algoritmo numérico que maximiza una función objetivo bajo ciertas condiciones. La función objetivo es la covarianza entre los vectores de *scores* de \mathbf{X} e \mathbf{Y} , mientras que la condición suele ser la ortogonalidad de estos vectores. Sin embargo, existen varios algoritmos de PLS y no todos usan la misma función objetivo con las mismas restricciones, por lo que no dan lugar a las mismas soluciones. Al margen de la función objetivo que se use, normalmente se distingue entre dos variantes de PLS, conocidas como PLS1 (cuando \mathbf{Y} es un vector) y PLS2 (cuando tanto \mathbf{X} como \mathbf{Y} son variables multidimensionales).

Para PLS2, que puede verse como una generalización de PLS1, se asume que tanto \mathbf{X} como \mathbf{Y} son matrices de tamaño $n \times m$ y $n \times q$ respectivamente. Cada fila de \mathbf{X} se corresponde con una observación que se representa con m características (variables predictoras). La matriz \mathbf{Y} por su parte recoge q propiedades de cada una de las n observaciones. Para facilitar la notación vamos a suponer que las columnas de \mathbf{X} e \mathbf{Y} están centradas en la media. El objetivo de la regresión

PLS2 es encontrar una relación lineal

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (8.1)$$

entre las variables de \mathbf{X} e \mathbf{Y} usando una matriz de coeficientes, \mathbf{B} , de tamaño $m \times q$ y una matriz de errores \mathbf{E} . En caso de PLS1, el problema se reduce a

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}_1 \quad (8.2)$$

con coeficientes \mathbf{b} y error \mathbf{e}_1 . En lugar de buscar la relación directamente, ambos, \mathbf{X} e \mathbf{Y} , son modelados por variables latentes de acuerdo a los siguientes modelos de regresión:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X \quad (8.3)$$

y

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y \quad (8.4)$$

con matrices de error \mathbf{E}_X y \mathbf{E}_Y . Las matrices \mathbf{T} y \mathbf{U} (matrices de *scores*) así como las matrices \mathbf{P} y \mathbf{Q} (matrices *loadings*) tienen a columnas, con $a \leq \min(m, q, n)$ siendo éste el número de componentes PLS. Los valores de \mathbf{T} son combinaciones lineales de los valores de \mathbf{X} por lo que se pueden considerar un buen resumen de \mathbf{X} . De la misma forma, \mathbf{U} es un buen resumen de \mathbf{Y} . En adelante, \mathbf{t}_j , \mathbf{u}_j , \mathbf{p}_j y \mathbf{q}_j ($j = 1, \dots, a$) serán usados para denotar la j -ésima columna de \mathbf{T} , \mathbf{U} , \mathbf{P} y \mathbf{Q} respectivamente. Además llamaremos *scores-x* y *scores-y* a los valores de \mathbf{T} y \mathbf{U} respectivamente puesto que estas matrices representan a los vectores de *scores* de \mathbf{X} y de \mathbf{Y} .

Los valores de \mathbf{T} y \mathbf{U} están conectados por la siguiente relación lineal:

$$\mathbf{u}_j = d_j \mathbf{t}_j + \mathbf{h}_j \quad (8.5)$$

siendo \mathbf{h}_j los residuos y d_j los parámetros de la regresión. Si, por ejemplo, la relación lineal entre \mathbf{u}_1 y \mathbf{t}_1 es fuerte (los elementos de \mathbf{h}_1 son pequeños), entonces los *scores-x* de la primera componente de PLS predicen bastante bien los *scores-y* y por tanto predicen también de forma adecuada los valores de \mathbf{Y} . No obstante, en la mayoría de las ocasiones se usa más de una componente para modelar \mathbf{Y} a partir de \mathbf{X} . La relación entre las matrices de *scores* queda por tanto de la siguiente forma

$$\mathbf{U} = \mathbf{TD} + \mathbf{H} \quad (8.6)$$

siendo \mathbf{D} una matriz diagonal con elementos d_1, d_2, \dots, d_a y \mathbf{H} la matriz de residuos (con \mathbf{h}_j nos referiremos a las columnas de \mathbf{H}). El nombre de PLS está motivado por la ecuación (8.6) ya que sólo parte de la información de \mathbf{X} y de \mathbf{Y} es usada para la regresión. Para PLS1, donde no se dispone de *scores-y*, la ecuación (8.6) se reduce a

$$\mathbf{y} = \mathbf{T}\mathbf{d} + \mathbf{h} \quad (8.7)$$

El objetivo de PLS2 es maximizar la covarianza entre los *scores-x* y los *scores-y*, mientras que para PLS1 se debe maximizar la covarianza entre los *scores-x* e \mathbf{y} . Usar la covarianza como un criterio para calcular variables latentes combina alta varianza de \mathbf{X} y alta correlación entre \mathbf{X} e \mathbf{Y} aunque esto depende en gran medida de cómo sea estimada la covarianza. En el caso clásico, la covarianza entre dos vectores \mathbf{t} y \mathbf{u} se calcula como la covarianza muestral, $\mathbf{t}^T \mathbf{u} / (n - 1)$, aunque también se pueden usar estimadores más robustos. Dado que el problema de maximización puede no ser único, se necesita una constante sobre los vectores de *scores* y ésta suele ser $\|\mathbf{t}\| = \|\mathbf{u}\| = 1$ (longitud 1). Los vectores de *scores* se obtienen a partir de la proyección de las matrices \mathbf{X} e \mathbf{Y} sobre los vectores de *loadings*. Aunque parece lógico usar aquí los vectores de *loadings* que forman las matrices \mathbf{P} y \mathbf{Q} (ecuaciones 8.3 y 8.4), por motivos técnicos usaremos otros vectores de *loadings*. Concretamente \mathbf{w} para \mathbf{X} y \mathbf{c} para \mathbf{Y}

$$\mathbf{t} = \mathbf{X}\mathbf{w}$$

$$\mathbf{u} = \mathbf{Y}\mathbf{c}$$

El problema de maximización con restricciones queda así

$$\text{cov}(\mathbf{X}\mathbf{w}, \mathbf{Y}\mathbf{c}) \rightarrow \max \text{ cumpliendo que } \begin{cases} \|\mathbf{t}\| = \|\mathbf{X}\mathbf{w}\| = 1 \\ \|\mathbf{u}\| = \|\mathbf{Y}\mathbf{c}\| = 1 \end{cases} \quad (8.8)$$

Nótese que la restricción puede ser que la longitud de los vectores de *scores* sea 1 (como se ha elegido aquí) o que los vectores de pesos, \mathbf{w} y \mathbf{c} sean 1. La solución del problema de maximización son los primeros vectores de *scores*, es decir, \mathbf{t}_1 y \mathbf{u}_1 . Para los siguientes vectores de *scores* se sigue el mismo criterio pero añadiendo restricciones adicionales. Normalmente, estas nuevas restricciones consisten en la ortogonalidad de los vectores de *scores* previos, esto es, $\mathbf{t}_j^T \mathbf{t}_l = 0$ y $\mathbf{u}_j^T \mathbf{u}_l = 0$ para $1 \leq j < l < a$. Una estrategia alternativa consiste en exigir la ortogonalidad de los vectores de *loadings* lo que lleva a vectores de *scores* no ortogonales y por tanto correlacionados.

Para encontrar la primera componente y dado que usamos la covarianza muestral, el problema de maximización se puede reescribir como la maximización de

$$\mathbf{t}^T \mathbf{u} = (\mathbf{X}\mathbf{w})^T \mathbf{Y}\mathbf{c} = \mathbf{w}^T \mathbf{X}^T \mathbf{Y}\mathbf{c} \rightarrow \max \quad (8.9)$$

bajo las mismas restricciones de longitud de los vectores igual a 1. Las soluciones para \mathbf{w} y \mathbf{c} se encuentran fácilmente por medio de la Descomposición en Valores Singulares (SVD) [Henry & Hofrichter, 1992] de $\mathbf{X}^T \mathbf{Y}$. De acuerdo a esto, de entre todas las posibles direcciones de \mathbf{w} y \mathbf{c} , el máximo para la ecuación (8.9) se obtiene para los vectores $\mathbf{w} = \mathbf{w}_1$ y $\mathbf{c} = \mathbf{c}_1$ correspondientes al mayor valor singular de $\mathbf{X}^T \mathbf{Y}$ [Höskuldsson, 1988].

8.3. Algoritmos PLS

Se han propuesto diferentes algoritmos para resolver el problema de maximización planteado en la ecuación (8.9). En esta sección se describirán los 2 algoritmos más usados. Ambos siguen la estrategia de exigir la ortogonalidad de los vectores de *scores*, dando lugar a vectores de *scores* no correlacionados.

8.3.1. NIPALS

El algoritmo NIPALS fue el primer algoritmo para resolver el problema de PLS y actualmente existen varias versiones del mismo con diferencias muy pequeñas entre unas y otras. Aquí se describe la versión más usada. Para calcular la primera componente PLS se procede como sigue:

1. Inicializar \mathbf{u}_1 con la primera columna de \mathbf{Y}
 2. $\mathbf{w}_1 = \frac{\mathbf{X}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}$
 3. $\mathbf{w}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}$
 4. $\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1$
 5. $\mathbf{c}_1 = \frac{\mathbf{Y}^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1}$
 6. $\mathbf{c}_1 = \frac{\mathbf{c}_1}{\|\mathbf{c}_1\|}$
 7. $\mathbf{u}_1^* = \mathbf{Y}\mathbf{c}_1$
-

$$8. \mathbf{u}_\Delta = \mathbf{u}_1^* - \mathbf{u}_1$$

$$9. \Delta \mathbf{u} = \mathbf{u}_\Delta^T \mathbf{u}_\Delta$$

10. Si $\Delta \mathbf{u} > \epsilon$, $\mathbf{u}_1 = \mathbf{u}_1^*$ e ir al paso 2; en otro caso, parar.

Considerando las ecuaciones 8.3 y 8.4, puede verse como en el paso 2 se buscan los coeficientes \mathbf{w}_1^j en el modelo de regresión $\mathbf{X} = \mathbf{u}_1^{j-1}(\mathbf{w}_1^j)^T + \mathbf{e}$ con un termino de error \mathbf{e} . El denominador $\mathbf{u}_1^T \mathbf{u}_1$ se muestra para que la regresión se vea claramente. Después de normalizar \mathbf{w} en el paso 3, \mathbf{X} es proyectada en el paso 4 para buscar sus vectores de *scores*. En el paso 5 hay una regresión por mínimos cuadrados de \mathbf{Y} sobre el primer vector de *scores* de \mathbf{X} . La normalización del paso 6 y la proyección del paso 7 actualizan el primer vector de *scores* de \mathbf{Y} . Este esquema de regresión alternativa de los datos de \mathbf{X} e \mathbf{Y} normalmente converge rápidamente.

No es fácil ver como este algoritmo soluciona el problema inicial (ecuación 8.8) de maximización de la covarianza entre los vectores de *scores* de \mathbf{X} e \mathbf{Y} . Pero si relacionamos las ecuaciones con el pseudocódigo anterior es más fácil verlo. Por ejemplo si nos situamos en el paso 2, para la iteración $j + 1$ obtenemos $\mathbf{w}_1^{j+1} = \frac{\mathbf{X}^T \mathbf{u}_1^j}{(\mathbf{u}_1^j)^T \mathbf{u}_1^j}$. Aquí podemos reemplazar \mathbf{u}_1^j por su valor (paso 7) y posteriormente reemplazamos \mathbf{c}_1 por su valor (pasos 5 y 6) y así sucesivamente hasta obtener:

$$\mathbf{w}_1^{j+1} = \mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X} \mathbf{w}_1^j \cdot m \quad (8.10)$$

donde m es una constante que depende de la norma de los diferentes vectores. La ecuación 8.10 es un problema de valor propio donde \mathbf{w}_1 es el vector propio $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$ para el mayor valor propio. De forma similar, se puede comprobar como $\mathbf{c}_1^{j+1} = \mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y} \mathbf{c}_1^j \cdot m$.

Para las siguiente componentes PLS, NIPALS requiere una reducción de \mathbf{X} e \mathbf{Y} por lo que el pseudocódigo anterior continua de la siguiente forma:

$$11. \mathbf{p}_1 = \frac{\mathbf{X}^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1}$$

$$12. \mathbf{q}_1 = \frac{\mathbf{Y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}$$

$$13. d_1 = \frac{\mathbf{u}_1^T \mathbf{t}_1}{\mathbf{t}_1^T \mathbf{t}_1}$$

$$14. \mathbf{X}_1 = \mathbf{X} - \mathbf{t}_1 \mathbf{p}_1^T \text{ y } \mathbf{Y}_1 = \mathbf{Y} - d_1 \mathbf{t}_1 \mathbf{c}_1^T$$

Las matrices \mathbf{X}_1 y \mathbf{Y}_1 resultantes son usadas para calcular las siguientes componentes PLS siguiendo el esquema de los pasos 1-10. Finalmente los coeficientes de la regresión de la ecuación (8.1), \mathbf{B} , que relacionan \mathbf{X} e \mathbf{Y} son obtenidos como

$$\mathbf{B} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{C}^T \quad (8.11)$$

Aquí las matrices \mathbf{W} , \mathbf{P} y \mathbf{C} se forman con los vectores \mathbf{w}_j , \mathbf{p}_j y \mathbf{c}_j ($j = 1, \dots, a$), siendo a el número de componentes PLS. Para PLS1, el algoritmo NIPALS se simplifica ya que no es necesario usar un algoritmo iterativo para extraer la primera componente.

8.3.2. SIMPLS

Propuesto por Jong en 1993, este algoritmo maximiza directamente el problema inicial (ecuación 8.8) bajo la restricción de ortogonalidad de los *scores-t* [de Jong, 1993]. La primera componente de PLS se calcula de forma idéntica a como lo hace el algoritmo NIPALS pero las siguientes se calculan de forma ligeramente diferente. La principal diferencia es que SIMPLS no reduce las matrices \mathbf{X} e \mathbf{Y} sino que la reducción es llevada a cabo sobre la matriz de covarianza, o más concretamente, sobre la matriz $\mathbf{S} = \mathbf{X}^T \mathbf{Y}$. A continuación se muestra la descripción del algoritmo en pseudocódigo

1. Inicializar $\mathbf{S}_0 = \mathbf{X}^T \mathbf{Y}$ y repetir los pasos 2 a 8 para $j = 1, \dots, a$
2. Si $j = 1$, $\mathbf{S}_j = \mathbf{S}_0$, sino $\mathbf{S}_j = \mathbf{S}_{j-1} - \mathbf{P}_{j-1}(\mathbf{P}_{j-1}^T \mathbf{P}_{j-1})^{-1} \mathbf{P}_{j-1}^T \mathbf{S}_{j-1}$
3. Calcular \mathbf{w}_j como el primer (izquierda) vector singular de \mathbf{S}_j
4. $\mathbf{w}_j = \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}$
5. $\mathbf{t}_j = \mathbf{X} \mathbf{w}_j$
6. $\mathbf{t}_j = \frac{\mathbf{t}_j}{\|\mathbf{t}_j\|}$
7. $\mathbf{p}_j = \mathbf{X}_j^T \mathbf{t}_j$
8. $\mathbf{P}_j = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{j-1}]$

Los pesos \mathbf{w}_j y los *scores* \mathbf{t}_j resultantes son almacenados como las columnas de las matrices \mathbf{W} y \mathbf{T} respectivamente. Nótese que la matriz \mathbf{W} difiere de la obtenida con otros algoritmos porque los pesos están relacionados directamente

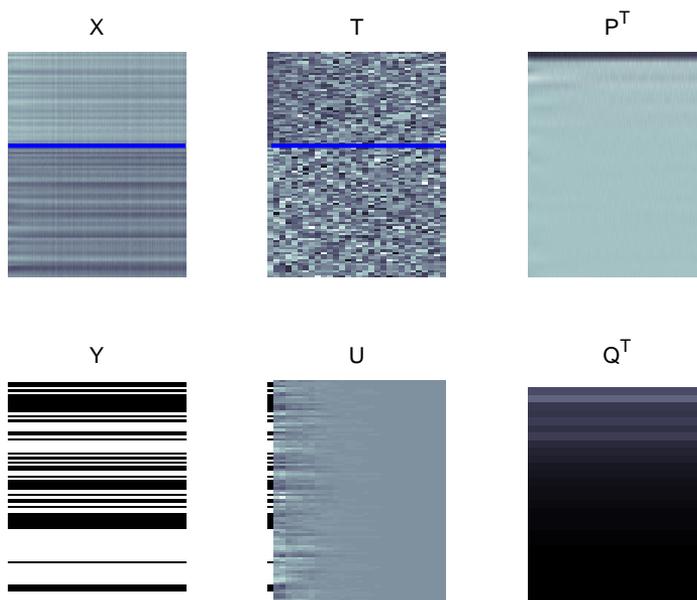


Figura 8.1: Matrices involucradas en el algoritmo PLS. \mathbf{X} agrupa una selección de vóxeles de 97 imágenes SPECT e \mathbf{Y} contiene las etiquetas (control o AD) asignadas a esas imágenes. \mathbf{T} y \mathbf{U} son las matrices de *scores* correspondientes a \mathbf{X} e \mathbf{Y} respectivamente. Análogamente, \mathbf{P} y \mathbf{Q} son las matrices de *loadings* de \mathbf{X} e \mathbf{Y} .

con \mathbf{X} y no con las matrices reducidas. El paso 2 hace que se cumpla la restricción de ortogonalidad de los *scores*- \mathbf{t}_j respecto a los vectores *scores* anteriores dado que la búsqueda se realiza en el complemento ortogonal de \mathbf{S}_{j-1} . El paso 3 maximiza directamente el problema inicial (ecuación 8.8). Los *scores* se obtienen en el paso 5 proyectando \mathbf{X} en la dirección óptima y los *loadings* son obtenidos en el paso 7 mediante mínimos cuadrados (ecuación 8.3).

Los coeficientes finales para la ecuación 8.1 son ahora

$$\mathbf{B} = \mathbf{W}\mathbf{T}^T\mathbf{Y} \quad (8.12)$$

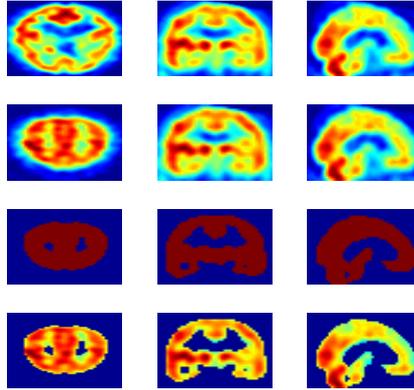


Figura 8.2: Procesado previo a la extracción de los vectores de *scores* de una imagen SPECT. *Fila superior*: Imagen original. *Segunda fila*: Imagen reducida. *Tercera fila*: Mascara de intensidad. *Fila inferior*: Imagen reducida una vez aplicada la mascara de intensidad

Para PLS1 el algoritmo se simplifica ligeramente. En este caso la ortogonalidad en el paso 2 se consigue usando la siguiente proyección cuando $j > 1$:

$$\mathbf{S}_j = \mathbf{S}_{j-1} - \mathbf{p}_{j-1}(\mathbf{p}_{j-1}^t \mathbf{p}_{j-1})^{-1} \mathbf{p}_{j-1}^T \mathbf{S}_{j-1} \quad (8.13)$$

8.4. Extracción de características

La extracción de los vectores de características, una vez ejecutado el algoritmo PLS, es trivial. Como se ha visto en la sección 8.2, un algoritmo PLS parte de dos conjuntos de datos, \mathbf{X} e \mathbf{Y} , y ajusta cada uno de ellos como un producto de dos matrices a las que se conoce como matriz de *loadings* y matriz de *scores* (ecuaciones 8.3 y 8.4). En nuestro caso, el conjunto \mathbf{X} es una matriz que contiene las imágenes tomográficas mientras que el conjunto \mathbf{Y} es un vector con las etiquetas (imagen normal o imagen patológica) correspondientes a estas imágenes. Se trata por tanto de la variante PLS1 de PLS.

Para construir \mathbf{X} a partir de las matrices tridimensionales correspondientes a las imágenes, en primer lugar es necesario reducir éstas a la mitad (en cada

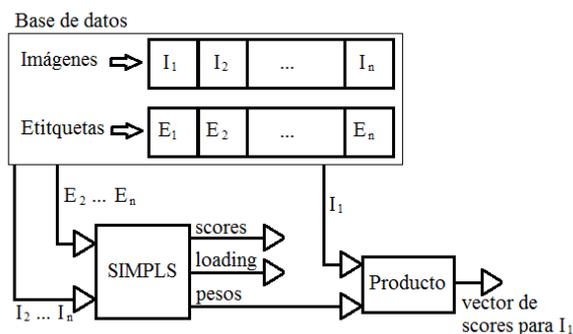


Figura 8.3: Esquema del funcionamiento del algoritmo de extracción de características basado en PLS. Se representa el proceso seguido para construir el vector de *scores* correspondiente a la primera imagen de la base de datos.

dimensión), pasando de imágenes con $69 \times 95 \times 79$ vóxeles a otras con $35 \times 48 \times 40$ vóxeles. De esta forma reducimos el tiempo de cómputo y mejoramos la efectividad del algoritmo al eliminar información redundante pues las marcas de Alzheimer en las imágenes tomográficas no están a nivel de vóxeles sino a nivel de estructuras superiores. Posteriormente se aplica una máscara de intensidad con el objetivo de seleccionar sólo aquellos vóxeles que forman parte del cerebro. Esta máscara se construye tomando los vóxeles con una intensidad superior al 50 % de la intensidad máxima. La figura 8.2 ilustra este proceso. Tras aplicar la máscara, los vóxeles resultantes de cada imagen se ponen en forma vectorial quedando el conjunto de todas las imágenes como una matriz bidimensional.

La figura 8.1 contiene una representación de las matrices involucradas en el algoritmo PLS. La matriz \mathbf{X} contiene los vóxeles de 97 imágenes SPECT (una imagen por fila) de la base de datos del hospital Virgen de las Nieves mientras que \mathbf{Y} es un vector con las 97 etiquetas asignadas a esas imágenes. Esta figura permite apreciar como la matriz de *scores*, \mathbf{T} , contiene la información relevante que nos permitirá separar las imágenes en clases, mientras que la matriz de *loadings*, \mathbf{P} , es mucho más homogénea.

En este se ha usado el algoritmo SIMPLS para llevar a cabo la descomposición en *scores* y *loadings* (ecuaciones (8.3) y (8.4)). Los parámetros de entrada del algoritmo (matrices \mathbf{X} e \mathbf{Y}) se construyen de la siguiente forma: La matriz \mathbf{X} está formada por las imágenes (una fila por imagen y una columna para cada vóxel) y la matriz \mathbf{Y} es en realidad un vector con las etiquetas de las imágenes

de \mathbf{X} . Es importante destacar que en la extracción de características se ha seguido una estrategia *leave-one-out* con el fin de evitar que la etiqueta de una imagen sea tomada en cuenta en la extracción de características para esa imagen. Por tanto, para cada imagen de la base de datos, se ha ejecutado el algoritmo con las imágenes restantes (\mathbf{X} contiene $n - 1$ imágenes e \mathbf{Y} las etiquetas correspondientes a las imágenes que hay en \mathbf{X}). En cada ejecución se obtiene, además de las matrices de *scores* y *loadings*, una matriz de pesos con la que se calcula el vector de *scores* para la imagen actual. Este vector de *scores* es usado usado como vector de características. La figura 8.3 muestra un esquema del proceso de extracción de características para una imagen de la base de datos.

CAPÍTULO

9

RESULTADOS EXPERIMENTALES

El rendimiento de los modelos expuestos en los dos capítulos anteriores ha sido evaluado mediante la implementación de varios sistemas CAD basados en ellos. Las tasas de precisión, sensibilidad y especificidad de los mismos son comparadas con enfoques anteriores como PCA [López *et al.*, 2009b] y VAF [Stoeckel *et al.*, 2001]. Para la realización de estos experimentos se han usado las tres bases de datos de imágenes funcionales descritas en el capítulo 3. En este capítulo se recogen los resultados obtenidos agrupados en 3 secciones, una para cada una de las bases de datos.

9.1. Descripción de los experimentos

Con el fin de evaluar el rendimiento de los métodos de extracción de características descritos en los capítulos 7 y 8 se han construido varios sistemas CAD con diferentes clasificadores estadísticos. Concretamente los clasificadores usados son:

- Clasificador de la media más cercana (NM)
- Clasificador lineal de Fisher (FLD)
- Clasificador bayesiano ingenuo (NB)
- Análisis discriminante lineal (LDA)
- Clasificador de los k vecinos más cercanos (KNN)
- Árboles de decisión (DT)
- Random Forests (RF)
- Maquinas de vectores soporte (SVM)

Para SVM se usaron dos *kernels*: lineal y función de base radial (RBF). Este último con valor del parámetro $\sigma = 5$.

En total, han sido implementados 5 métodos de extracción de características diferentes: VAF y PCA usados como *baseline*; GMM con las dos aproximaciones descritas en la sección 7.5, esto es, usando una media de controles (GMM₁) o la diferencia entre controles y imágenes patológicas (GMM₂) para delimitar las ROIs; y el método basado en PLS (capítulo 8). Dado que se han analizado todas las combinaciones posibles entre estos 5 métodos y los 9 clasificadores, en total hay 45 sistemas CAD diferentes.

Para los sistemas basados en el modelo de mezcla de Gaussianas, cada uno de los experimentos se realizó 64 veces usando vectores de características de tamaño 1..,64 en cada simulación. Tal y como se vio en la sección 7.5, un modelo con 64 Gaussianas produce vectores de características de 64 componentes, no obstante, es posible reducir estos vectores usando sólo las Gaussianas de mayor altura. De forma similar, los sistemas basados en PLS se repitieron n veces variando el tamaño de los vectores de características, cuyo tamaño máximo depende del tamaño de la base de datos ($n = \text{tamaño}BD - 1$). El sistema basado en VAF y el clasificador FLD no ha sido simulado debido a problemas de tiempo de

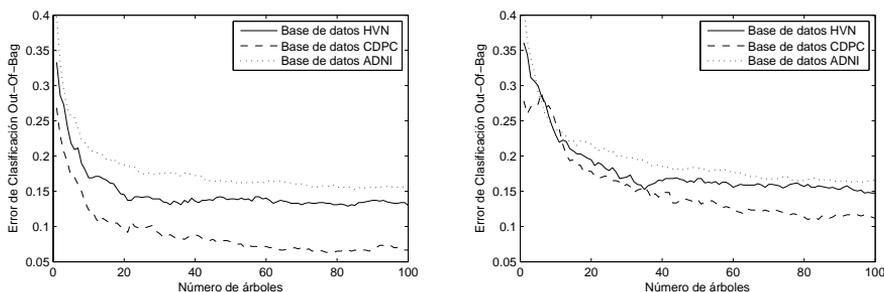


Figura 9.1: Error de clasificación Out-Of-Bag obtenido con el clasificador Random Forests en función del número de árboles usados. Se muestra el error usando tanto el modelo basado en GMM (izquierda) como el modelo basado en PLS (derecha)

cómputo motivados por el gran tamaño de los vectores de características que genera este modelo.

La validación se ha llevado a cabo mediante las técnicas leave-one-out y k -fold (con $k = 5$ y $k = 10$ que suelen ser los valores típicos [Hastie *et al.*, 2003]). Todos los sistemas han sido entrenados y validados con las tres bases de datos de imágenes descritas en el capítulo 3. Las siguientes secciones muestran los resultados obtenidos con cada una de las bases de datos usadas.

9.2. Ajuste de parámetros de los clasificadores

Los parámetros de los clasificadores se ajustan durante el proceso de entrenamiento, a partir de los datos de ejemplo. Sin embargo, algunos parámetros, como es el caso del número de árboles de un clasificador Random Forests, deben ser ajustados manualmente.

Para ajustar el número de árboles de Random Forests, hemos calculado el error de clasificación Out-Of-Bag, que representa la suma del error de clasificación que obtiene cada árbol del bosque al clasificar los ejemplos que no se usan en su construcción (recordemos que cada árbol se construye sólo con un subconjunto de los ejemplos del conjunto de entrenamiento). En la figura 9.1 se representa el error de clasificación Out-Of-Bag en función del número de árboles usados. Obsérvese como a partir de 50 árboles aproximadamente, el error permanece estable por lo que no tiene sentido seguir aumentando el número de

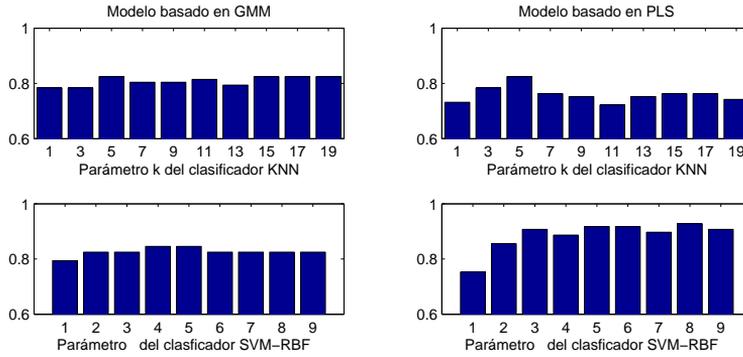


Figura 9.2: Precisión de los clasificadores KNN (fila superior) y SVM-RBF (fila inferior) en función del parámetro k y σ respectivamente. Estos resultados se han obtenido mediante validación 5-fold usando las imágenes de la base de datos HVN y los enfoques GMM (columna izquierda) y PLS (columna derecha)

árboles ya que no se conseguirán mejores resultados y en cambio se aumentarán los tiempos de ejecución del algoritmo.

Otros parámetros, como el valor k en el clasificador de los k vecinos más cercanos o el parámetro σ del clasificador SVM-RBF se calcularon empíricamente. La figura 9.2 muestra un estudio del comportamiento de ambos clasificadores en función de estos parámetros. Para este estudio se usó la base de datos HVN y los vectores de características completos (sin truncar). La precisión representada se obtiene mediante validación 5-fold. Los valores de k se han elegido de forma que sea un número impar para evitar empates en el número de vecinos más cercanos de cada clase.

9.3. Resultados para la base de datos HVN

En esta base de datos las imágenes están etiquetadas como controles, posible Alzheimer (AD1), probable Alzheimer (AD2) y Alzheimer seguro (AD3). Si bien, dado el reto al que se enfrenta este trabajo, los sistemas aquí diseñados distinguen en controles y enfermos de Alzheimer (cualquiera que sea su grado).

En la tabla 9.1 se muestra un resumen de la precisión obtenida por los sistemas basados en extracción de características mediante GMM y PLS comparadas con las obtenidos por enfoques anteriores como VAF y PCA. Este resumen mues-

tra la tasa de acierto de los sistemas estimada mediante validación leave-one-out.

Tabla 9.1: Precisión obtenida por los diferentes sistemas CAD construidos para la base de datos HVN. Se muestra el valor máximo alcanzado por cada sistema, independientemente del número de características.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	85.57 %	85.57 %	83.51 %	88.66 %	85.57 %
FLD	-	82.47 %	80.41 %	84.54 %	87.63 %
NB	86.60 %	88.66 %	88.66 %	89.69 %	88.66 %
LDA	-	85.57 %	81.44 %	89.69 %	95.88 %
KNN	80.41 %	82.47 %	84.54 %	85.57 %	84.54 %
DT	83.51 %	79.38 %	82.47 %	90.72 %	86.60 %
RF	84.54 %	88.66 %	89.69 %	89.69 %	87.63 %
SVM (linear)	83.51 %	88.66 %	85.57 %	89.69 %	96.91 %
SVM (RBF)	57.73 %	89.69 %	89.69 %	91.75 %	94.85 %

En los siguientes apartados se muestra una tabla que recoge la precisión, sensibilidad y especificidad para cada uno de los clasificadores usados.

Tabla 9.2: Clasificador de la media más cercana

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	85.57 %	85.57 %	83.51 %	88.66 %	85.57 %
Precisión media	85.57 %	83.68 %	80.27 %	85.07 %	60.22 %
Desv. precisión	-	0.79	6.82	1.27	12.18
Sensibilidad máxima	90.24 %	90.24 %	87.80 %	100.00 %	90.24 %
Especificidad máxima	82.14 %	82.14 %	80.36 %	80.36 %	82.14 %
Validación 5-fold:					
Precisión máxima	73.20 %	85.57 %	82.47 %	89.69 %	80.41 %
Precisión media	73.20 %	69.92 %	68.98 %	79.03 %	56.78 %
Desv. precisión	-	7.50	8.55	7.67	7.48
Sensibilidad máxima	68.29 %	90.24 %	87.80 %	97.56 %	87.80 %
Especificidad máxima	76.79 %	82.14 %	78.57 %	83.93 %	75.00 %
Validación 10-fold:					
Precisión máxima	84.54 %	85.57 %	82.47 %	90.72 %	77.32 %
Precisión media	84.54 %	70.81 %	68.35 %	82.28 %	56.23 %
Desv. precisión	-	6.38	6.81	5.19	7.78
Sensibilidad máxima	82.93 %	95.12 %	85.37 %	97.56 %	78.05 %
Especificidad máxima	85.71 %	78.57 %	80.36 %	85.71 %	76.79 %

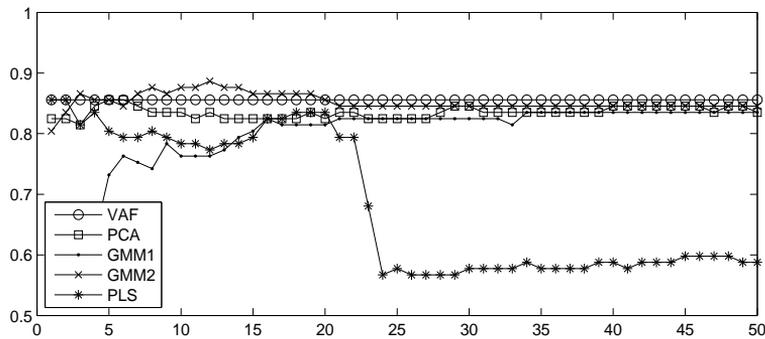


Figura 9.3: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.3: Clasificador lineal de Fisher

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	-	82.47 %	80.41 %	84.54 %	87.63 %
Precisión media	-	67.18 %	65.64 %	70.49 %	68.25 %
Desv. precisión	-	6.55	5.89	6.39	12.99
Sensibilidad máxima	-	87.80 %	90.24 %	92.68 %	92.68 %
Especificidad máxima	-	78.57 %	73.21 %	78.57 %	83.93 %
Validación 5-fold:					
Precisión máxima	-	69.07 %	70.10 %	80.41 %	85.57 %
Precisión media	-	54.82 %	56.94 %	56.39 %	54.38 %
Desv. precisión	-	5.41	5.16	8.63	7.15
Sensibilidad máxima	-	65.85 %	60.98 %	82.93 %	90.24 %
Especificidad máxima	-	71.43 %	76.79 %	78.57 %	82.14 %
Validación 10-fold:					
Precisión máxima	-	82.47 %	73.20 %	80.41 %	82.47 %
Precisión media	-	57.89 %	60.24 %	59.76 %	58.59 %
Desv. precisión	-	5.84	5.45	6.48	7.58
Sensibilidad máxima	-	87.80 %	75.61 %	82.93 %	87.80 %
Especificidad máxima	-	78.57 %	71.43 %	78.57 %	78.57 %

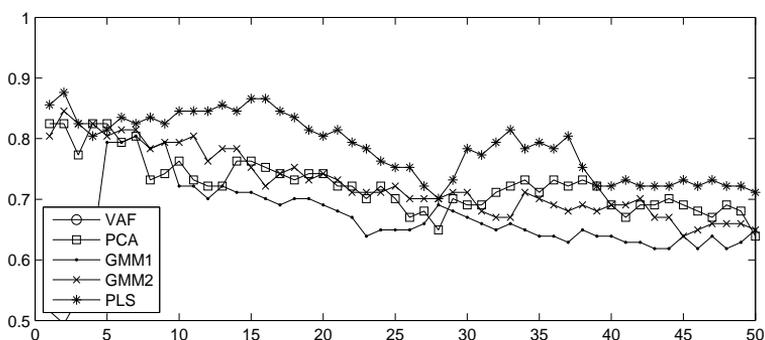


Figura 9.4: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.4: Clasificador bayesiano ingenuo

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	86.60 %	88.66 %	88.66 %	89.69 %	88.66 %
Precisión media	86.60 %	74.48 %	84.60 %	87.32 %	73.72 %
Desv. precisión	-	5.09	7.35	1.51	11.75
Sensibilidad máxima	90.24 %	87.80 %	90.24 %	95.12 %	90.24 %
Especificidad máxima	83.93 %	89.29 %	87.50 %	85.71 %	87.50 %
Validación 5-fold:					
Precisión máxima	86.60 %	84.54 %	89.69 %	89.69 %	88.66 %
Precisión media	86.60 %	73.75 %	84.46 %	87.08 %	74.32 %
Desv. precisión	-	5.16	7.38	1.55	11.34
Sensibilidad máxima	90.24 %	82.93 %	90.24 %	92.68 %	78.05 %
Especificidad máxima	83.93 %	85.71 %	89.29 %	87.50 %	96.43 %
Validación 10-fold:					
Precisión máxima	87.63 %	87.63 %	88.66 %	89.69 %	89.69 %
Precisión media	87.63 %	74.16 %	84.37 %	87.27 %	74.27 %
Desv. precisión	-	5.46	7.11	1.55	11.69
Sensibilidad máxima	90.24 %	87.80 %	90.24 %	95.12 %	92.68 %
Especificidad máxima	85.71 %	87.50 %	87.50 %	85.71 %	87.50 %

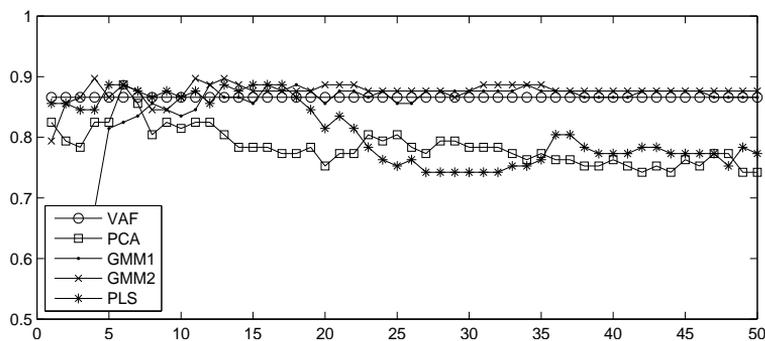


Figura 9.5: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.5: Análisis Discriminante Lineal

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	-	85.57 %	81.44 %	89.69 %	95.88 %
Precisión media	-	70.53 %	72.42 %	82.33 %	85.96 %
Dev. precisión	-	11.10	7.39	2.72	6.26
Sensibilidad máxima	-	92.68 %	90.24 %	92.68 %	100.00 %
Especificidad máxima	-	80.36 %	75.00 %	87.50 %	92.86 %
Validación 5-fold:					
Precisión máxima	-	86.60 %	82.47 %	89.69 %	95.88 %
Precisión media	-	71.56 %	70.68 %	80.69 %	83.15 %
Dev. precisión	-	9.38	8.11	5.18	8.53
Sensibilidad máxima	-	95.12 %	90.24 %	95.12 %	100.00 %
Especificidad máxima	-	80.36 %	76.79 %	85.71 %	92.86 %
Validación 10-fold:					
Precisión máxima	-	85.57 %	84.54 %	90.72 %	93.81 %
Precisión media	-	71.83 %	72.13 %	81.39 %	84.21 %
Dev. precisión	-	9.96	7.74	3.83	6.89
Sensibilidad máxima	-	92.68 %	90.24 %	92.68 %	95.12 %
Especificidad máxima	-	80.36 %	80.36 %	89.29 %	92.86 %

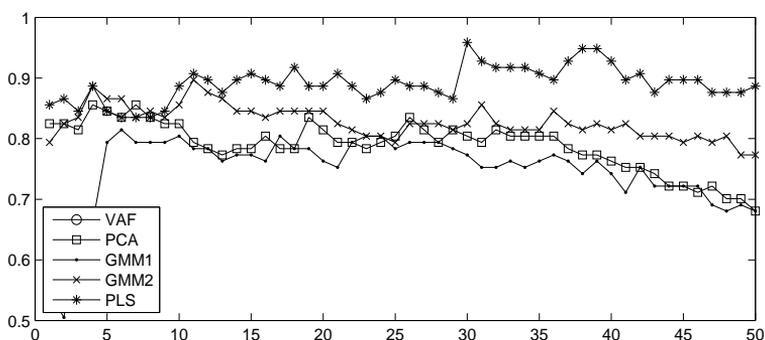


Figura 9.6: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.6: Clasificador de los k vecinos más cercanos

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	80.41 %	82.47 %	84.54 %	85.57 %	84.54 %
Precisión media	80.41 %	79.22 %	80.07 %	83.47 %	63.39 %
Desv. precisión	-	2.26	6.36	0.80	11.13
Sensibilidad máxima	87.80 %	85.37 %	92.68 %	85.37 %	87.80 %
Especificidad máxima	75.00 %	80.36 %	78.57 %	85.71 %	82.14 %
Validación 5-fold:					
Precisión máxima	79.38 %	84.54 %	86.60 %	88.66 %	84.54 %
Precisión media	79.38 %	78.70 %	80.57 %	84.81 %	62.59 %
Desv. precisión	-	2.52	6.45	1.54	10.89
Sensibilidad máxima	90.24 %	92.68 %	92.68 %	90.24 %	92.68 %
Especificidad máxima	71.43 %	78.57 %	82.14 %	87.50 %	78.57 %
Validación 10-fold:					
Precisión máxima	80.41 %	85.57 %	86.60 %	86.60 %	84.54 %
Precisión media	80.41 %	79.06 %	80.35 %	84.10 %	63.08 %
Desv. precisión	-	2.34	6.65	1.11	10.66
Sensibilidad máxima	87.80 %	87.80 %	95.12 %	87.80 %	90.24 %
Especificidad máxima	75.00 %	83.93 %	80.36 %	85.71 %	80.36 %

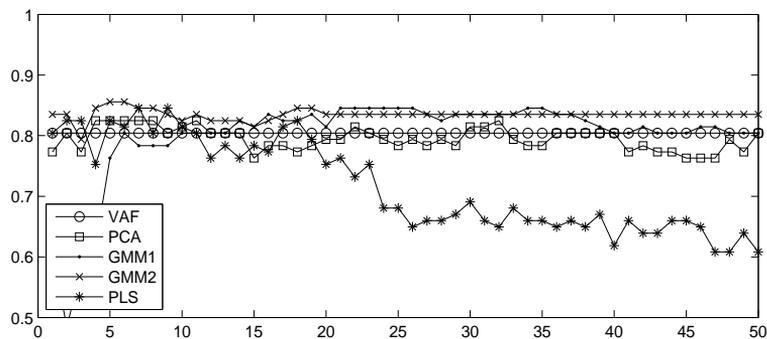


Figura 9.7: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.7: Árboles de Decisión

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	83.51 %	79.38 %	82.47 %	90.72 %	86.60 %
Precisión media	83.51 %	72.20 %	79.04 %	88.16 %	80.44 %
Desv. precisión	-	2.29	6.42	3.07	2.87
Sensibilidad máxima	80.49 %	73.17 %	85.37 %	87.80 %	82.93 %
Especificidad máxima	85.71 %	83.93 %	80.36 %	92.86 %	89.29 %
Validación 5-fold:					
Precisión máxima	84.54 %	83.51 %	86.60 %	93.81 %	85.57 %
Precisión media	84.54 %	74.08 %	78.04 %	87.26 %	78.58 %
Desv. precisión	-	3.73	6.49	3.38	3.23
Sensibilidad máxima	85.37 %	78.05 %	87.80 %	90.24 %	82.93 %
Especificidad máxima	83.93 %	87.50 %	85.71 %	96.43 %	87.50 %
Validación 10-fold:					
Precisión máxima	84.54 %	81.44 %	86.60 %	93.81 %	86.60 %
Precisión media	84.54 %	73.41 %	78.88 %	87.81 %	78.98 %
Desv. precisión	-	3.56	7.62	3.63	3.25
Sensibilidad máxima	82.93 %	75.61 %	85.37 %	87.80 %	82.93 %
Especificidad máxima	85.71 %	85.71 %	87.50 %	98.21 %	89.29 %

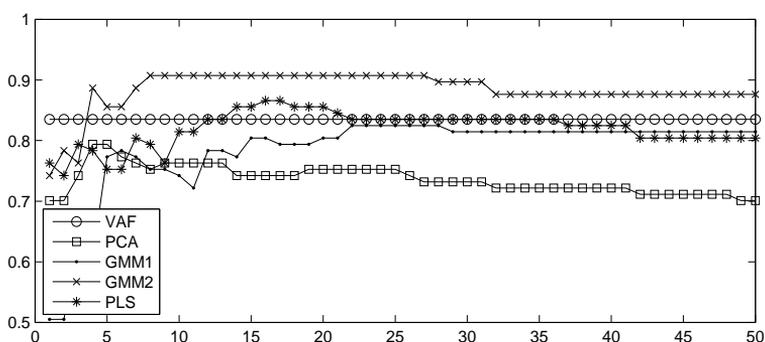


Figura 9.8: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.8: Random Forests

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	84.54 %	88.66 %	89.69 %	89.69 %	87.63 %
Precisión media	84.54 %	79.37 %	82.99 %	86.66 %	82.49 %
Desv. precisión	-	4.44	7.00	2.65	2.71
Sensibilidad máxima	80.49 %	78.05 %	85.37 %	87.80 %	82.93 %
Especificidad máxima	87.50 %	96.43 %	92.86 %	91.07 %	91.07 %
Validación 5-fold:					
Precisión máxima	85.57 %	87.63 %	90.72 %	90.72 %	89.69 %
Precisión media	85.57 %	77.05 %	82.59 %	86.50 %	81.56 %
Desv. precisión	-	5.06	7.60	2.75	4.15
Sensibilidad máxima	82.93 %	82.93 %	87.80 %	87.80 %	85.37 %
Especificidad máxima	87.50 %	91.07 %	92.86 %	92.86 %	92.86 %
Validación 10-fold:					
Precisión máxima	85.57 %	87.63 %	88.66 %	89.69 %	91.75 %
Precisión media	85.57 %	78.77 %	82.94 %	86.95 %	82.22 %
Desv. precisión	-	4.90	7.24	2.62	3.58
Sensibilidad máxima	85.37 %	82.93 %	87.80 %	87.80 %	87.80 %
Especificidad máxima	85.71 %	91.07 %	89.29 %	91.07 %	94.64 %

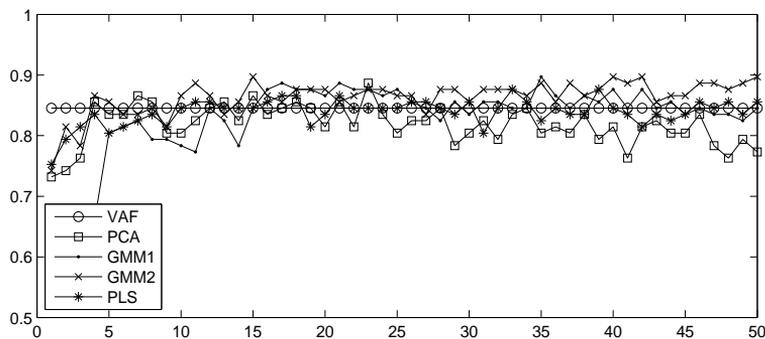


Figura 9.9: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.9: Maquinas de Vectores Soporte (lineal)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	83.51 %	88.66 %	85.57 %	89.69 %	96.91 %
Precisión media	83.51 %	69.60 %	78.93 %	83.89 %	92.75 %
Desv. precisión	-	10.37	5.97	2.41	3.25
Sensibilidad máxima	80.49 %	90.24 %	80.49 %	85.37 %	95.12 %
Especificidad máxima	85.71 %	87.50 %	89.29 %	92.86 %	98.21 %
Validación 5-fold:					
Precisión máxima	82.47 %	88.66 %	86.60 %	89.69 %	95.88 %
Precisión media	82.47 %	69.42 %	79.40 %	84.78 %	89.12 %
Desv. precisión	-	10.16	6.01	2.79	3.16
Sensibilidad máxima	82.93 %	85.37 %	85.37 %	87.80 %	95.12 %
Especificidad máxima	82.14 %	91.07 %	87.50 %	91.07 %	96.43 %
Validación 10-fold:					
Precisión máxima	84.54 %	88.66 %	86.60 %	91.75 %	97.94 %
Precisión media	84.54 %	70.00 %	79.12 %	84.18 %	90.77 %
Desv. precisión	-	10.71	5.78	2.60	3.46
Sensibilidad máxima	85.37 %	90.24 %	82.93 %	92.68 %	100.00 %
Especificidad máxima	83.93 %	87.50 %	89.29 %	91.07 %	96.43 %

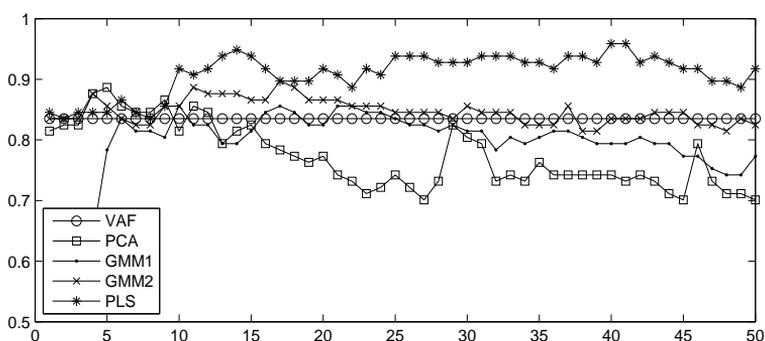


Figura 9.10: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.10: Maquinas de Vectores Soporte (RBF)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	57.73 %	89.69 %	89.69 %	91.75 %	94.85 %
Precisión media	57.73 %	73.39 %	85.76 %	88.10 %	84.05 %
Desv. precisión	-	11.00	6.33	1.85	4.01
Sensibilidad máxima	0.00 %	87.80 %	85.37 %	90.24 %	92.68 %
Especificidad máxima	100.00 %	91.07 %	92.86 %	92.86 %	96.43 %
Validación 5-fold:					
Precisión máxima	57.73 %	91.75 %	90.72 %	90.72 %	94.85 %
Precisión media	57.73 %	71.84 %	85.02 %	87.92 %	82.75 %
Desv. precisión	-	10.40	6.19	1.67	5.39
Sensibilidad máxima	0.00 %	92.68 %	87.80 %	87.80 %	90.24 %
Especificidad máxima	100.00 %	91.07 %	92.86 %	92.86 %	98.21 %
Validación 10-fold:					
Precisión máxima	57.73 %	90.72 %	91.75 %	91.75 %	94.85 %
Precisión media	57.73 %	72.66 %	85.28 %	87.98 %	83.79 %
Desv. precisión	-	10.70	6.44	1.88	4.78
Sensibilidad máxima	0.00 %	90.24 %	87.80 %	90.24 %	92.68 %
Especificidad máxima	100.00 %	91.07 %	94.64 %	92.86 %	96.43 %

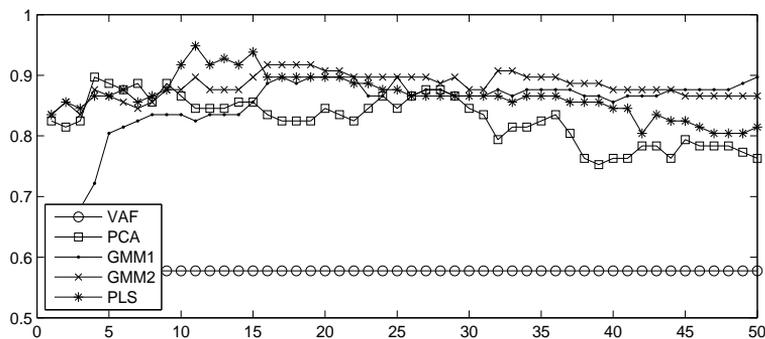


Figura 9.11: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

9.3.1. Análisis de los resultados

A priori, es lógico pensar que la precisión obtenida por los experimentos que hacen uso de esta base de datos es inferior a la de las otras bases de datos formadas por imágenes PET. Como se vio en la sección 2.3, la técnica SPECT usada para obtener las imágenes de la base de datos HVN proporciona imágenes con menor resolución que la técnica PET.

La precisión de un clasificador depende fundamentalmente de dos factores, a saber, el tamaño del conjunto de entrenamiento y la dimensión de los vectores de características. A mayor tamaño del conjunto de entrenamiento, mejor será el ajuste de los parámetros de la función de clasificación (ecuación (5.2)) y esto redundará en una mayor precisión. En este sentido, la complejidad del clasificador resulta determinante para obtener tasas de acierto elevadas, ya que un clasificador muy complejo tendrá un mayor número de parámetros y por tanto, necesitará más ejemplos que un clasificador más simple para ajustar satisfactoriamente todos los parámetros.

De forma similar, el tamaño de los vectores de características determina la dimensión del problema y, a mayor dimensión, más difícil es ajustar la función de clasificación que separa correctamente los ejemplos. Si el tamaño de los vectores de características es grande, el número de parámetros de la función de clasificación, inherente al tipo de clasificador, juega un papel determinante. Obsérvese, como para el método VAF, que produce vectores de características muy grandes, la precisión es mayor con clasificadores simples como NM o NB. Sin embargo otros clasificadores más complejos como SVM, que requieren estimar un mayor número de parámetros para ajustar la función de clasificación, obtienen tasas de acierto más bajas. Especialmente baja resulta la tasa de acierto obtenida por SVM con *kernel* RBF. El motivo hay que buscarlo en la dificultad que tienen los clasificadores no lineales, como SVM-RBF, para ajustar sus parámetros cuando la dimensión del problema es elevada. Por otro lado, con el resto métodos diferentes a VAF, donde el tamaño de los vectores de características es sensiblemente inferior, SVM obtiene mejores resultados que otros clasificadores más simples, sobre todo en su variante no lineal, es decir, con *kernel* RBF.

Respecto a los métodos presentados en esta Tesis Doctoral, es importante notar que ambos superan, por lo general, a enfoques anteriores basados en VAF o en PCA. La mejor tasa de acierto para esta base de datos, 96.91 %, se obtiene mediante el método basado en PLS y un clasificador SVM lineal, sin embargo, el modelo denominado GMM₂ es más estable y mantiene tasas de acierto que rondan el 90 % con la mayoría de los clasificadores.

9.4. Resultados para la base de datos CDPC

Las imágenes de la base de datos del Centro de Diagnóstico PET Cartuja sólo tienen dos posibles etiquetas, controles y enfermos de Alzheimer, los sistemas CAD han sido entrenados para distinguir entre estos dos estados.

En la tabla 9.11 se muestra un resumen de las tasas de acierto obtenidas por los sistemas basados en extracción de características mediante GMM y PLS comparadas con las obtenidas por enfoques anteriores como VAF y PCA. Este resumen muestra la precisión de los sistemas estimada mediante validación leave-one-out.

Tabla 9.11: Precisión obtenida por los diferentes sistemas CAD construidos para la base de datos CDPC. Se muestra el valor máximo alcanzado por cada sistema, independientemente del número de características.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	86.67 %	90.00 %	75.00 %	88.33 %	95.00 %
FLD	-	98.33 %	98.33 %	98.33 %	98.33 %
NB	88.33 %	95.00 %	73.33 %	93.33 %	95.00 %
LDA	-	98.33 %	96.67 %	96.67 %	88.33 %
KNN	91.67 %	93.33 %	88.33 %	95.00 %	91.67 %
DT	85.00 %	86.67 %	86.67 %	90.00 %	88.33 %
RF	93.33 %	93.33 %	93.33 %	95.00 %	91.67 %
SVM (linear)	96.67 %	96.67 %	96.67 %	96.67 %	91.67 %
SVM (RBF)	70.00 %	96.67 %	96.67 %	98.33 %	91.67 %

En los siguientes apartados se muestra una tabla que recoge la precisión, sensibilidad y especificidad para cada uno de los clasificadores usados.

Tabla 9.12: Clasificador de la media más cercana

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	86.67 %	90.00 %	75.00 %	88.33 %	95.00 %
Precisión media	86.67 %	87.71 %	70.63 %	87.66 %	79.51 %
Desv. precisión	-	4.93	3.67	1.18	16.41
Sensibilidad máxima	94.44 %	100.00 %	88.89 %	94.44 %	100.00 %
Especificidad máxima	83.33 %	85.71 %	69.05 %	85.71 %	92.86 %
Validación 5-fold:					
Precisión máxima	76.67 %	83.33 %	78.33 %	81.67 %	78.33 %
Precisión media	76.67 %	75.99 %	68.91 %	77.73 %	66.93 %
Desv. precisión	-	3.75	4.86	1.27	7.98
Sensibilidad máxima	94.44 %	100.00 %	94.44 %	100.00 %	94.44 %
Especificidad máxima	69.05 %	76.19 %	71.43 %	73.81 %	71.43 %
Validación 10-fold:					
Precisión máxima	78.33 %	81.67 %	73.33 %	86.67 %	86.67 %
Precisión media	78.33 %	77.12 %	68.88 %	83.70 %	73.79 %
Desv. precisión	-	3.24	3.76	1.17	12.55
Sensibilidad máxima	88.89 %	94.44 %	83.33 %	94.44 %	94.44 %
Especificidad máxima	73.81 %	76.19 %	69.05 %	83.33 %	83.33 %

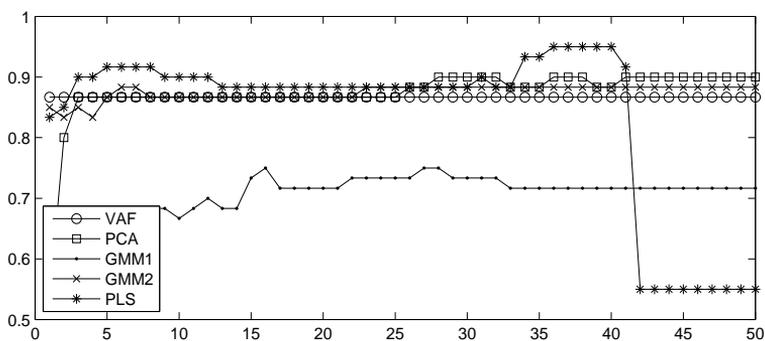


Figura 9.12: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.13: Clasificador lineal de Fisher

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	-	98.33 %	98.33 %	98.33 %	98.33 %
Precisión media	-	96.55 %	92.50 %	95.63 %	87.07 %
Desv. precisión	-	6.05	10.63	4.45	13.82
Sensibilidad máxima	-	100.00 %	100.00 %	100.00 %	100.00 %
Especificidad máxima	-	97.62 %	97.62 %	97.62 %	97.62 %
Validación 5-fold:					
Precisión máxima	-	88.33 %	86.67 %	83.33 %	81.67 %
Precisión media	-	79.75 %	76.54 %	75.99 %	72.67 %
Desv. precisión	-	6.79	8.01	3.25	9.37
Sensibilidad máxima	-	100.00 %	100.00 %	94.44 %	100.00 %
Especificidad máxima	-	83.33 %	80.95 %	78.57 %	73.81 %
Validación 10-fold:					
Precisión máxima	-	95.00 %	93.33 %	95.00 %	91.67 %
Precisión media	-	90.76 %	87.34 %	89.01 %	81.78 %
Desv. precisión	-	5.50	9.35	3.80	12.97
Sensibilidad máxima	-	100.00 %	100.00 %	100.00 %	94.44 %
Especificidad máxima	-	92.86 %	90.48 %	92.86 %	90.48 %

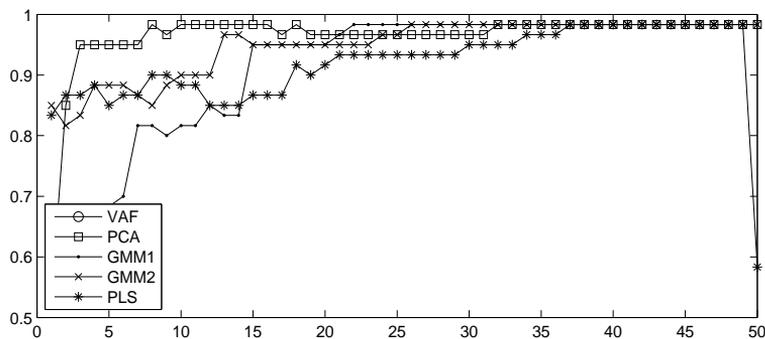


Figura 9.13: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.14: Clasificador bayesiano ingenuo

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	88.33 %	95.00 %	73.33 %	93.33 %	95.00 %
Precisión media	88.33 %	89.86 %	68.75 %	90.76 %	88.99 %
Desv. precisión	-	5.52	5.38	1.70	2.68
Sensibilidad máxima	83.33 %	88.89 %	83.33 %	94.44 %	94.44 %
Especificidad máxima	90.48 %	97.62 %	69.05 %	92.86 %	95.24 %
Validación 5-fold:					
Precisión máxima	88.33 %	95.00 %	76.67 %	95.00 %	95.00 %
Precisión media	88.33 %	88.70 %	70.18 %	90.52 %	89.40 %
Desv. precisión	-	5.82	5.96	2.20	2.55
Sensibilidad máxima	83.33 %	88.89 %	83.33 %	94.44 %	94.44 %
Especificidad máxima	90.48 %	97.62 %	73.81 %	95.24 %	95.24 %
Validación 10-fold:					
Precisión máxima	90.00 %	95.00 %	76.67 %	95.00 %	95.00 %
Precisión media	90.00 %	89.35 %	69.38 %	90.49 %	89.40 %
Desv. precisión	-	4.74	5.72	1.94	2.77
Sensibilidad máxima	83.33 %	88.89 %	83.33 %	94.44 %	94.44 %
Especificidad máxima	92.86 %	97.62 %	73.81 %	95.24 %	95.24 %

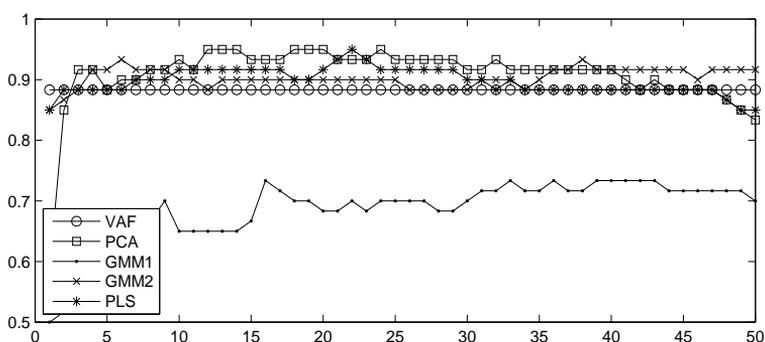


Figura 9.14: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.15: Análisis Discriminante Lineal

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	-	98.33 %	96.67 %	96.67 %	88.33 %
Precisión media	-	91.75 %	83.68 %	90.96 %	79.81 %
Desv. precisión	-	8.98	12.66	4.48	6.94
Sensibilidad máxima	-	100.00 %	100.00 %	100.00 %	83.33 %
Especificidad máxima	-	97.62 %	95.24 %	95.24 %	90.48 %
Validación 5-fold:					
Precisión máxima	-	100.00 %	96.67 %	95.00 %	88.33 %
Precisión media	-	91.05 %	82.83 %	86.99 %	77.83 %
Desv. precisión	-	10.03	11.05	5.95	8.79
Sensibilidad máxima	-	100.00 %	100.00 %	100.00 %	83.33 %
Especificidad máxima	-	100.00 %	95.24 %	92.86 %	90.48 %
Validación 10-fold:					
Precisión máxima	-	98.33 %	96.67 %	96.67 %	90.00 %
Precisión media	-	91.60 %	83.17 %	89.07 %	78.56 %
Desv. precisión	-	8.35	11.34	6.50	8.73
Sensibilidad máxima	-	100.00 %	94.44 %	94.44 %	83.33 %
Especificidad máxima	-	97.62 %	97.62 %	97.62 %	92.86 %

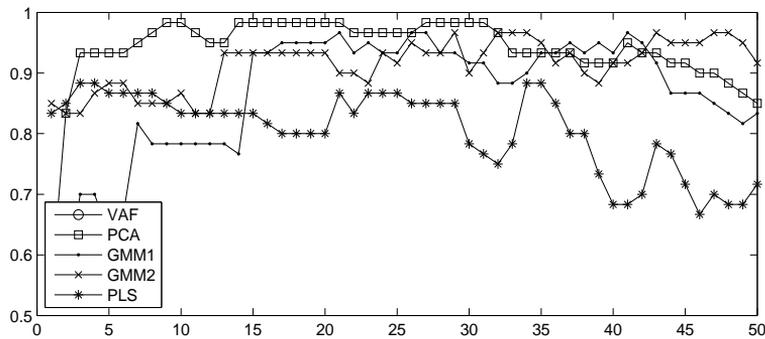


Figura 9.15: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.16: Clasificador de los k vecinos más cercanos

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	91.67 %	93.33 %	88.33 %	95.00 %	91.67 %
Precisión media	91.67 %	90.03 %	82.40 %	93.10 %	74.71 %
Desv. precisión	-	4.23	5.44	3.15	9.43
Sensibilidad máxima	100.00 %	94.44 %	88.89 %	94.44 %	77.78 %
Especificidad máxima	88.10 %	92.86 %	88.10 %	95.24 %	97.62 %
Validación 5-fold:					
Precisión máxima	93.33 %	93.33 %	88.33 %	96.67 %	93.33 %
Precisión media	93.33 %	89.01 %	80.86 %	92.84 %	74.05 %
Desv. precisión	-	4.20	6.22	3.21	10.01
Sensibilidad máxima	100.00 %	94.44 %	88.89 %	94.44 %	83.33 %
Especificidad máxima	90.48 %	92.86 %	88.10 %	97.62 %	97.62 %
Validación 10-fold:					
Precisión máxima	93.33 %	93.33 %	88.33 %	96.67 %	91.67 %
Precisión media	93.33 %	89.97 %	81.61 %	92.73 %	74.71 %
Desv. precisión	-	3.56	6.51	3.38	9.72
Sensibilidad máxima	100.00 %	94.44 %	94.44 %	94.44 %	77.78 %
Especificidad máxima	90.48 %	92.86 %	85.71 %	97.62 %	97.62 %

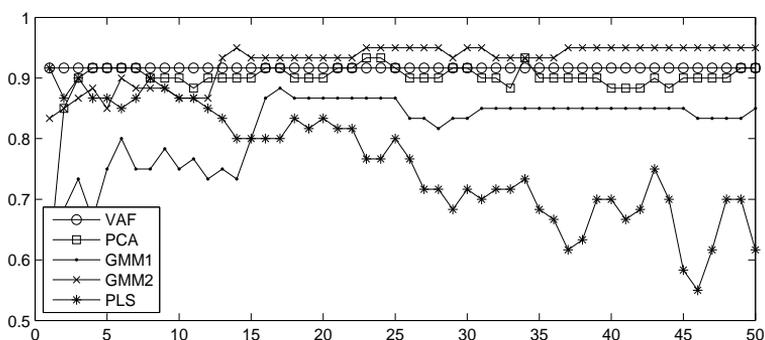


Figura 9.16: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.17: Árboles de Decisión

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	85.00 %	86.67 %	86.67 %	90.00 %	88.33 %
Precisión media	85.00 %	85.71 %	81.07 %	86.22 %	82.36 %
Desv. precisión	-	3.74	6.53	2.17	2.05
Sensibilidad máxima	77.78 %	66.67 %	77.78 %	88.89 %	77.78 %
Especificidad máxima	88.10 %	95.24 %	90.48 %	90.48 %	92.86 %
Validación 5-fold:					
Precisión máxima	90.00 %	91.67 %	91.67 %	96.67 %	90.00 %
Precisión media	90.00 %	83.05 %	79.48 %	85.57 %	83.48 %
Desv. precisión	-	6.37	7.86	4.26	3.05
Sensibilidad máxima	88.89 %	83.33 %	77.78 %	100.00 %	77.78 %
Especificidad máxima	90.48 %	95.24 %	97.62 %	95.24 %	95.24 %
Validación 10-fold:					
Precisión máxima	85.00 %	91.67 %	90.00 %	91.67 %	88.33 %
Precisión media	85.00 %	84.86 %	79.77 %	85.39 %	83.07 %
Desv. precisión	-	5.17	8.16	3.42	2.41
Sensibilidad máxima	83.33 %	77.78 %	83.33 %	88.89 %	77.78 %
Especificidad máxima	85.71 %	97.62 %	92.86 %	92.86 %	92.86 %

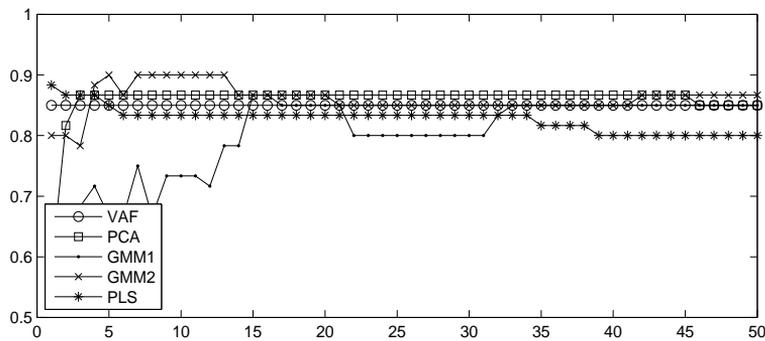


Figura 9.17: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.18: Random Forests

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	93.33 %	93.33 %	93.33 %	95.00 %	91.67 %
Precisión media	93.33 %	87.32 %	86.15 %	91.67 %	87.84 %
Desv. precisión	-	7.25	9.21	2.77	2.50
Sensibilidad máxima	83.33 %	83.33 %	83.33 %	88.89 %	77.78 %
Especificidad máxima	97.62 %	97.62 %	97.62 %	97.62 %	97.62 %
Validación 5-fold:					
Precisión máxima	90.00 %	95.00 %	95.00 %	96.67 %	91.67 %
Precisión media	90.00 %	86.67 %	84.45 %	91.09 %	86.70 %
Desv. precisión	-	7.27	9.17	3.49	2.99
Sensibilidad máxima	72.22 %	83.33 %	94.44 %	94.44 %	77.78 %
Especificidad máxima	97.62 %	100.00 %	95.24 %	97.62 %	97.62 %
Validación 10-fold:					
Precisión máxima	91.67 %	95.00 %	93.33 %	96.67 %	91.67 %
Precisión media	91.67 %	86.89 %	85.10 %	92.19 %	87.21 %
Desv. precisión	-	7.23	8.58	2.82	2.71
Sensibilidad máxima	83.33 %	83.33 %	83.33 %	94.44 %	77.78 %
Especificidad máxima	95.24 %	100.00 %	97.62 %	97.62 %	97.62 %

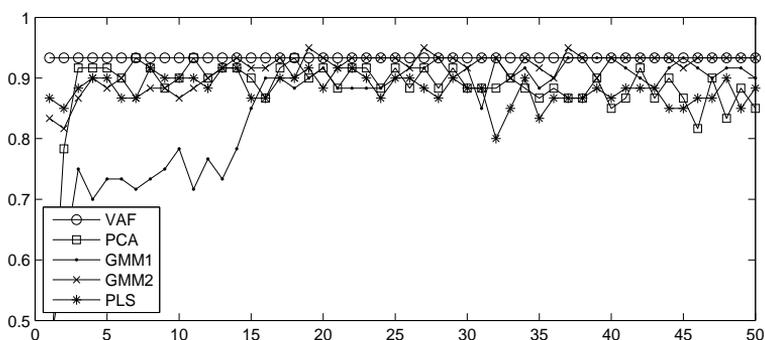


Figura 9.18: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.19: Maquinas de Vectores Soporte (lineal)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	96.67 %	96.67 %	96.67 %	96.67 %	91.67 %
Precisión media	96.67 %	89.12 %	92.37 %	94.66 %	80.20 %
Desv. precisión	-	8.04	6.74	3.63	5.84
Sensibilidad máxima	94.44 %	94.44 %	94.44 %	94.44 %	77.78 %
Especificidad máxima	97.62 %	97.62 %	97.62 %	97.62 %	97.62 %
Validación 5-fold:					
Precisión máxima	96.67 %	98.33 %	98.33 %	98.33 %	90.00 %
Precisión media	96.67 %	87.68 %	91.85 %	94.30 %	78.68 %
Desv. precisión	-	8.72	7.26	4.18	5.89
Sensibilidad máxima	94.44 %	100.00 %	100.00 %	100.00 %	72.22 %
Especificidad máxima	97.62 %	97.62 %	97.62 %	97.62 %	97.62 %
Validación 10-fold:					
Precisión máxima	96.67 %	98.33 %	96.67 %	98.33 %	91.67 %
Precisión media	96.67 %	89.01 %	91.80 %	94.56 %	79.25 %
Desv. precisión	-	7.94	7.23	3.84	5.68
Sensibilidad máxima	94.44 %	100.00 %	94.44 %	100.00 %	77.78 %
Especificidad máxima	97.62 %	97.62 %	97.62 %	97.62 %	97.62 %

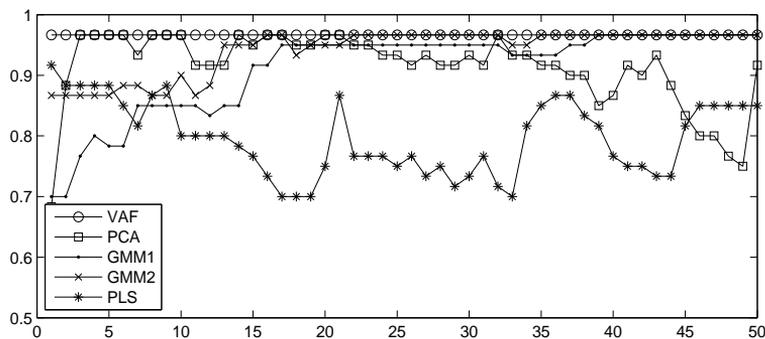


Figura 9.19: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.20: Maquinas de Vectores Soporte (RBF)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	70.00 %	96.67 %	96.67 %	98.33 %	91.67 %
Precisión media	70.00 %	83.64 %	89.48 %	93.65 %	88.07 %
Desv. precisión	-	10.47	9.51	2.78	1.65
Sensibilidad máxima	0.00 %	94.44 %	94.44 %	100.00 %	77.78 %
Especificidad máxima	100.00 %	97.62 %	97.62 %	97.62 %	97.62 %
Validación 5-fold:					
Precisión máxima	70.00 %	96.67 %	96.67 %	98.33 %	93.33 %
Precisión media	70.00 %	83.02 %	88.20 %	93.46 %	85.95 %
Desv. precisión	-	10.37	9.46	2.88	3.23
Sensibilidad máxima	0.00 %	94.44 %	94.44 %	100.00 %	83.33 %
Especificidad máxima	100.00 %	97.62 %	97.62 %	97.62 %	97.62 %
Validación 10-fold:					
Precisión máxima	70.00 %	96.67 %	96.67 %	96.67 %	93.33 %
Precisión media	70.00 %	83.45 %	89.19 %	93.39 %	87.59 %
Desv. precisión	-	10.45	9.09	2.91	2.69
Sensibilidad máxima	0.00 %	94.44 %	94.44 %	100.00 %	77.78 %
Especificidad máxima	100.00 %	97.62 %	97.62 %	95.24 %	100.00 %

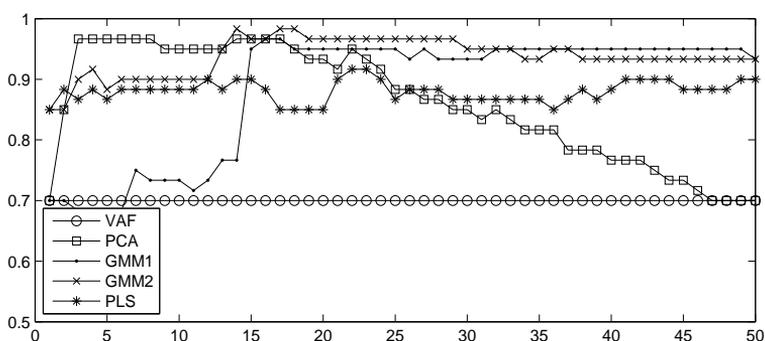


Figura 9.20: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

9.4.1. Análisis de los resultados

Los resultados obtenidos con esta base de datos son sensiblemente mejores que los obtenidos con la base de datos HVN. Esto es debido principalmente a la composición de esta base de datos. Mientras que la base de datos HVN contiene un elevado porcentaje de imágenes procedentes de pacientes dudosos (el 54 % de las imágenes patológicas están etiquetadas como “posible AD”), las imágenes de la base de datos CDPC corresponden a casos claros de enfermos de AD y controles, por lo que la clasificación es más fácil. Además, las imágenes obtenidas mediante la tomografía PET tienen mayor resolución que las obtenidas mediante SPECT y reflejan con mayor fidelidad las diferencias en la actividad cerebral entre personas sanas y enfermos de AD.

En este caso, el pico máximo de precisión, 98.33 %, se alcanza tanto con el enfoque PCA como con los sistemas basados en GMM y en PLS. A pesar de que PLS supone una mejora respecto a PCA en tanto en cuanto que tiene en cuenta las etiquetas de las imágenes para extraer los vectores de *scores*, PCA obtiene resultados más estables cuando lo combinamos con diferentes clasificadores. Esto se debe a que la selección de los *scores* basada solamente en la varianza de las imágenes (como hace PCA) permite separar satisfactoriamente las imágenes de esta base de datos. El hecho de tener en cuenta también la varianza de las etiquetas (como hace PLS) sólo aporta mayor complejidad a la extracción de características.

Por otro lado, resulta destacable el buen comportamiento que tiene el clasificador FLD. Este clasificador alcanza la máxima tasa de acierto con todos los métodos salvo con VAF para el que no se han obtenido resultados por problemas de cómputo (la ejecución del algoritmo requiere ingentes cantidades de memoria). Este hecho evidencia de nuevo que, con bases de datos pequeñas, la complejidad del clasificador, medida como el número de parámetros de la función de clasificación, juega un papel fundamental. Téngase en cuenta que esta base de datos dispone sólo de 60 imágenes por lo que los conjuntos de entrenamiento y test usados para la validación son pequeños.

Al igual que con la base de datos anterior, el método denominado GMM₂ es el que obtiene mejores resultados, superando el 95 % de precisión con la mayoría de los clasificadores. Además, tal y como muestran las gráficas con los resultados parciales para los diferentes clasificadores (de fig. 9.12 a fig. 9.20), este método es el que aporta resultados de clasificación más estables independientemente del tamaño de los vectores de características.

9.5. Resultados para la base de datos ADNI

Esta base de datos es la más completa de las tres en la medida en que dispone de más imágenes y de más etiquetas para clasificar esas imágenes. Por ello, hemos llevado a cabo dos experimentos distintos. En primer lugar hemos tenido en cuenta sólo los pacientes con Alzheimer y los controles, construyendo sistemas CAD que permiten determinar cuando una persona sufre AD. En segundo lugar hemos aprovechado la riqueza de la base de datos, al incluir pacientes con deterioro cognitiva leve (MCI), para construir sistemas CAD más exigentes que sean capaces incluso de determinar qué casos de MCI derivarán en AD y cuáles no. Para ello entrenamos los sistemas CAD de modo que separen, por un lado, controles y MCI no convertidos y, por otro lado, MCI convertidos (pacientes en los que el deterioro cognitivo derivó en AD con el paso del tiempo) y AD.

9.5.1. Controles frente a Alzheimer

En este caso se han usado 192 imágenes de la base de datos ADNI, las correspondientes a controles y pacientes con AD (véase la sección 3.3.2).

Tabla 9.21: Precisión obtenida por los diferentes sistemas CAD construidos para una selección de 192 imágenes de la base de datos ADNI (97 controles y 95 pacientes AD). Se muestra el valor máximo alcanzado por cada sistema, independientemente del número de características.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	81.25 %	80.21 %	80.21 %	85.94 %	83.33 %
FLD	-	81.25 %	72.92 %	79.69 %	82.29 %
NB	80.73 %	76.04 %	82.29 %	88.02 %	85.42 %
LDA	-	84.90 %	80.73 %	86.46 %	86.46 %
KNN	76.04 %	81.25 %	78.12 %	86.98 %	83.85 %
DT	75.00 %	73.44 %	77.08 %	84.38 %	82.29 %
RF	82.81 %	82.81 %	84.38 %	90.62 %	86.98 %
SVM (linear)	80.21 %	87.50 %	82.81 %	87.50 %	85.94 %
SVM (RBF)	50.52 %	85.94 %	84.90 %	89.58 %	85.42 %

En los siguientes apartados se muestra una tabla que recoge la precisión, sensibilidad y especificidad para cada uno de los clasificadores usados.

Tabla 9.22: Clasificador de la media más cercana

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	81.25 %	80.21 %	80.21 %	85.94 %	83.33 %
Precisión media	81.25 %	78.99 %	74.96 %	84.47 %	57.72 %
Desv. precisión	-	1.13	5.27	2.25	13.11
Sensibilidad máxima	89.69 %	87.63 %	89.69 %	92.78 %	90.72 %
Especificidad máxima	72.63 %	72.63 %	70.53 %	78.95 %	75.79 %
Validación 5-fold:					
Precisión máxima	57.81 %	71.88 %	66.67 %	83.85 %	76.56 %
Precisión media	57.81 %	58.66 %	56.94 %	67.89 %	51.51 %
Desv. precisión	-	6.19	5.18	10.50	5.58
Sensibilidad máxima	63.92 %	77.32 %	69.07 %	89.69 %	86.60 %
Especificidad máxima	51.58 %	66.32 %	64.21 %	77.89 %	66.32 %
Validación 10-fold:					
Precisión máxima	61.98 %	66.67 %	67.71 %	73.96 %	64.06 %
Precisión media	61.98 %	54.98 %	55.73 %	55.75 %	50.61 %
Desv. precisión	-	4.06	5.08	8.86	4.01
Sensibilidad máxima	65.98 %	67.01 %	65.98 %	73.20 %	73.20 %
Especificidad máxima	57.89 %	66.32 %	69.47 %	74.74 %	54.74 %

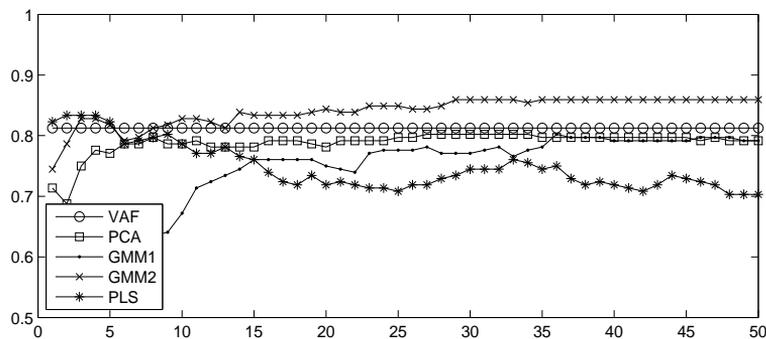


Figura 9.21: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.23: Clasificador lineal de Fisher

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	-	81.25 %	72.92 %	79.69 %	82.29 %
Precisión media	-	61.95 %	66.91 %	72.99 %	59.15 %
Desv. precisión	-	9.51	2.57	2.95	9.57
Sensibilidad máxima	-	85.57 %	80.41 %	86.60 %	85.57 %
Especificidad máxima	-	76.84 %	65.26 %	72.63 %	78.95 %
Validación 5-fold:					
Precisión máxima	-	71.88 %	66.15 %	66.67 %	79.17 %
Precisión media	-	52.09 %	53.86 %	54.92 %	51.75 %
Desv. precisión	-	4.32	4.61	5.44	5.20
Sensibilidad máxima	-	73.20 %	62.89 %	74.23 %	83.51 %
Especificidad máxima	-	70.53 %	69.47 %	58.95 %	74.74 %
Validación 10-fold:					
Precisión máxima	-	63.54 %	66.67 %	58.33 %	64.06 %
Precisión media	-	51.90 %	52.36 %	50.94 %	50.84 %
Desv. precisión	-	3.93	3.48	3.99	3.98
Sensibilidad máxima	-	59.79 %	67.01 %	60.82 %	67.01 %
Especificidad máxima	-	67.37 %	66.32 %	55.79 %	61.05 %

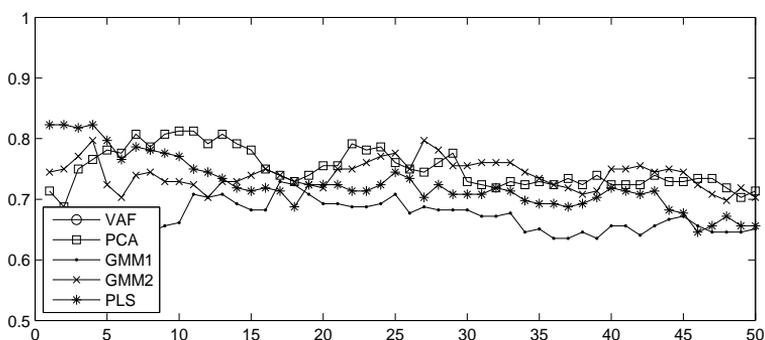


Figura 9.22: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.24: Clasificador bayesiano ingenuo

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	80.73 %	76.04 %	82.29 %	88.02 %	85.42 %
Precisión media	80.73 %	66.03 %	77.34 %	84.28 %	74.07 %
Desv. precisión	-	4.08	3.78	2.19	5.54
Sensibilidad máxima	89.69 %	83.51 %	87.63 %	90.72 %	89.69 %
Especificidad máxima	71.58 %	68.42 %	76.84 %	85.26 %	81.05 %
Validación 5-fold:					
Precisión máxima	79.69 %	75.52 %	82.81 %	88.02 %	84.38 %
Precisión media	79.69 %	66.01 %	77.28 %	84.29 %	73.43 %
Desv. precisión	-	4.46	3.52	2.25	5.30
Sensibilidad máxima	87.63 %	80.41 %	88.66 %	90.72 %	88.66 %
Especificidad máxima	71.58 %	70.53 %	76.84 %	85.26 %	80.00 %
Validación 10-fold:					
Precisión máxima	80.73 %	76.04 %	82.29 %	88.02 %	84.38 %
Precisión media	80.73 %	65.88 %	77.28 %	84.33 %	73.95 %
Desv. precisión	-	4.39	3.91	2.15	5.55
Sensibilidad máxima	89.69 %	82.47 %	87.63 %	90.72 %	87.63 %
Especificidad máxima	71.58 %	69.47 %	76.84 %	85.26 %	81.05 %

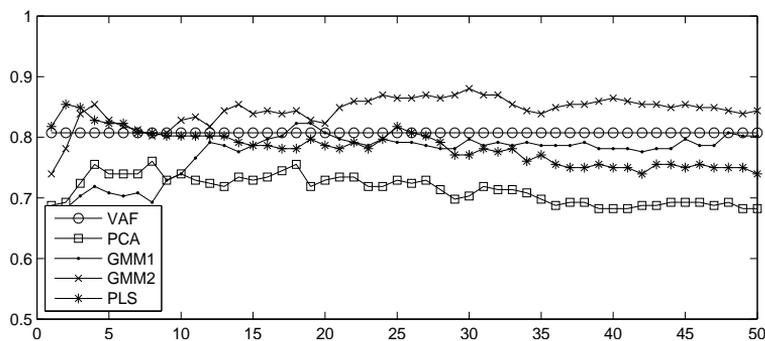


Figura 9.23: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.25: Análisis Discriminante Lineal

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	-	84.90 %	80.73 %	86.46 %	86.46 %
Precisión media	-	75.86 %	75.37 %	83.51 %	77.55 %
Desv. precisión	-	7.70	4.67	1.91	4.06
Sensibilidad máxima	-	88.66 %	86.60 %	88.66 %	92.78 %
Especificidad máxima	-	81.05 %	74.74 %	84.21 %	80.00 %
Validación 5-fold:					
Precisión máxima	-	85.94 %	80.73 %	85.94 %	85.94 %
Precisión media	-	75.86 %	74.41 %	82.37 %	75.99 %
Desv. precisión	-	7.12	4.36	2.30	5.44
Sensibilidad máxima	-	87.63 %	83.51 %	88.66 %	91.75 %
Especificidad máxima	-	84.21 %	77.89 %	83.16 %	80.00 %
Validación 10-fold:					
Precisión máxima	-	85.42 %	81.25 %	86.98 %	86.46 %
Precisión media	-	76.14 %	74.86 %	82.73 %	76.79 %
Desv. precisión	-	6.91	4.75	2.21	4.77
Sensibilidad máxima	-	87.63 %	88.66 %	89.69 %	91.75 %
Especificidad máxima	-	83.16 %	73.68 %	84.21 %	81.05 %

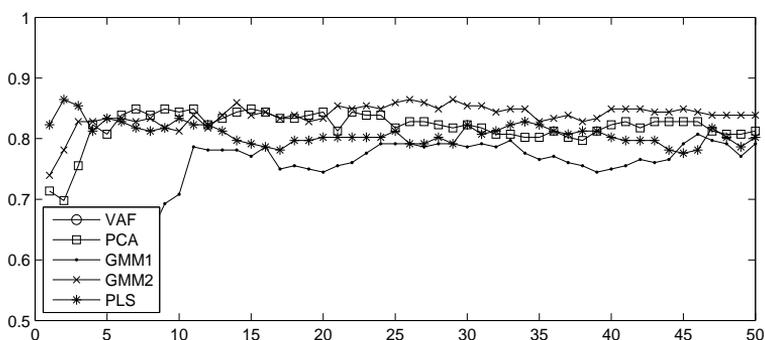


Figura 9.24: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.26: Clasificador de los k vecinos más cercanos

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	76.04 %	81.25 %	78.12 %	86.98 %	83.85 %
Precisión media	76.04 %	73.14 %	72.99 %	84.81 %	53.96 %
Desv. precisión	-	2.67	4.33	2.54	11.37
Sensibilidad máxima	88.66 %	89.69 %	89.69 %	92.78 %	84.54 %
Especificidad máxima	63.16 %	72.63 %	66.32 %	81.05 %	83.16 %
Validación 5-fold:					
Precisión máxima	76.56 %	81.25 %	79.17 %	88.02 %	83.85 %
Precisión media	76.56 %	73.10 %	72.90 %	84.29 %	54.38 %
Desv. precisión	-	2.67	4.49	2.71	10.75
Sensibilidad máxima	89.69 %	87.63 %	92.78 %	93.81 %	84.54 %
Especificidad máxima	63.16 %	74.74 %	65.26 %	82.11 %	83.16 %
Validación 10-fold:					
Precisión máxima	77.08 %	81.77 %	78.65 %	86.98 %	85.94 %
Precisión media	77.08 %	73.19 %	72.99 %	84.54 %	54.28 %
Desv. precisión	-	2.69	4.58	2.46	11.16
Sensibilidad máxima	86.60 %	85.57 %	91.75 %	93.81 %	86.60 %
Especificidad máxima	67.37 %	77.89 %	65.26 %	80.00 %	85.26 %

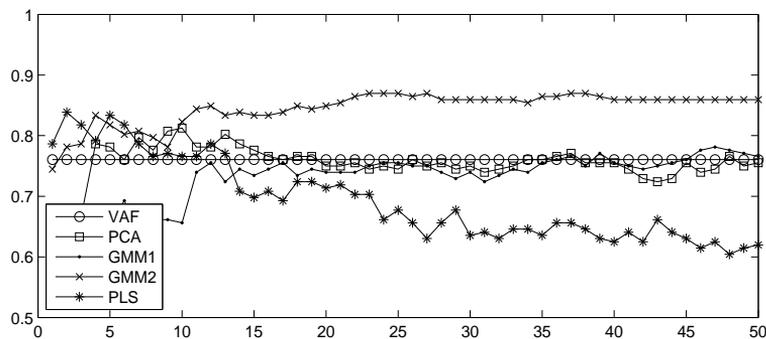


Figura 9.25: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.27: Árboles de Decisión

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	75.00 %	73.44 %	77.08 %	84.38 %	82.29 %
Precisión media	75.00 %	69.81 %	71.40 %	76.74 %	78.87 %
Desv. precisión	-	2.33	5.39	2.40	1.29
Sensibilidad máxima	73.20 %	76.29 %	77.32 %	85.57 %	78.35 %
Especificidad máxima	76.84 %	70.53 %	76.84 %	83.16 %	86.32 %
Validación 5-fold:					
Precisión máxima	76.04 %	79.17 %	78.12 %	83.33 %	82.29 %
Precisión media	76.04 %	67.73 %	70.55 %	77.28 %	74.54 %
Desv. precisión	-	3.55	4.69	2.65	3.38
Sensibilidad máxima	76.29 %	79.38 %	81.44 %	80.41 %	84.54 %
Especificidad máxima	75.79 %	78.95 %	74.74 %	86.32 %	80.00 %
Validación 10-fold:					
Precisión máxima	70.31 %	75.52 %	76.04 %	83.33 %	81.77 %
Precisión media	70.31 %	67.83 %	70.02 %	76.91 %	75.25 %
Desv. precisión	-	3.19	5.08	2.72	2.82
Sensibilidad máxima	69.07 %	74.23 %	81.44 %	82.47 %	81.44 %
Especificidad máxima	71.58 %	76.84 %	70.53 %	84.21 %	82.11 %

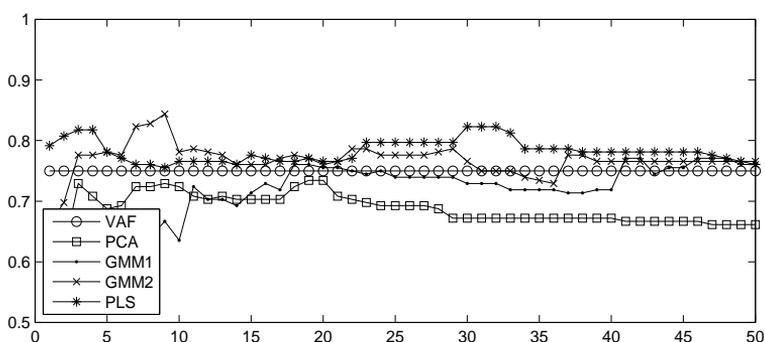


Figura 9.26: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.28: Random Forests

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	82.81 %	82.81 %	84.38 %	90.62 %	86.98 %
Precisión media	82.81 %	72.00 %	78.29 %	86.38 %	81.65 %
Desv. precisión	-	5.08	6.17	3.73	2.42
Sensibilidad máxima	86.60 %	83.51 %	85.57 %	88.66 %	88.66 %
Especificidad máxima	78.95 %	82.11 %	83.16 %	92.63 %	85.26 %
Validación 5-fold:					
Precisión máxima	83.33 %	82.81 %	85.94 %	90.62 %	86.46 %
Precisión media	83.33 %	70.21 %	78.62 %	86.17 %	80.50 %
Desv. precisión	-	5.79	6.39	3.58	2.71
Sensibilidad máxima	85.57 %	81.44 %	88.66 %	87.63 %	86.60 %
Especificidad máxima	81.05 %	84.21 %	83.16 %	93.68 %	86.32 %
Validación 10-fold:					
Precisión máxima	81.25 %	80.73 %	85.94 %	91.15 %	86.46 %
Precisión media	81.25 %	71.59 %	79.04 %	86.09 %	81.07 %
Desv. precisión	-	5.22	5.92	3.88	2.49
Sensibilidad máxima	85.57 %	83.51 %	86.60 %	90.72 %	88.66 %
Especificidad máxima	76.84 %	77.89 %	85.26 %	91.58 %	84.21 %

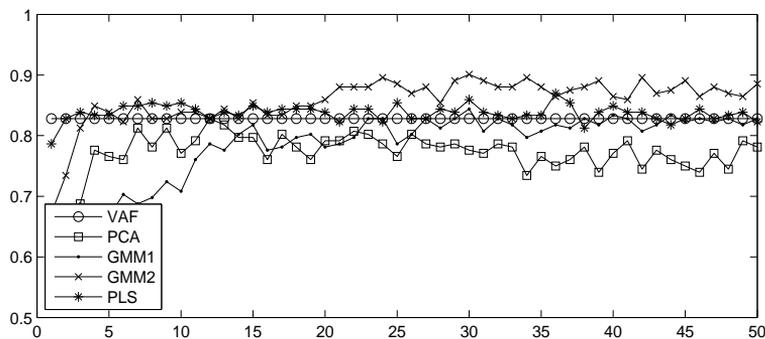


Figura 9.27: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.29: Maquinas de Vectores Soporte (lineal)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	80.21 %	87.50 %	82.81 %	87.50 %	85.94 %
Precisión media	80.21 %	73.43 %	76.51 %	84.62 %	76.40 %
Desv. precisión	-	14.01	5.44	2.13	3.70
Sensibilidad máxima	80.41 %	87.63 %	85.57 %	90.72 %	89.69 %
Especificidad máxima	80.00 %	87.37 %	80.00 %	84.21 %	82.11 %
Validación 5-fold:					
Precisión máxima	80.21 %	86.98 %	82.81 %	87.50 %	87.50 %
Precisión media	80.21 %	72.33 %	76.25 %	84.26 %	77.03 %
Desv. precisión	-	10.16	5.75	2.35	4.14
Sensibilidad máxima	81.44 %	85.57 %	83.51 %	86.60 %	90.72 %
Especificidad máxima	78.95 %	88.42 %	82.11 %	88.42 %	84.21 %
Validación 10-fold:					
Precisión máxima	81.77 %	86.98 %	82.29 %	88.54 %	85.94 %
Precisión media	81.77 %	72.27 %	76.39 %	84.62 %	81.25 %
Desv. precisión	-	11.81	5.52	2.34	2.52
Sensibilidad máxima	82.47 %	85.57 %	85.57 %	89.69 %	88.66 %
Especificidad máxima	81.05 %	88.42 %	78.95 %	87.37 %	83.16 %

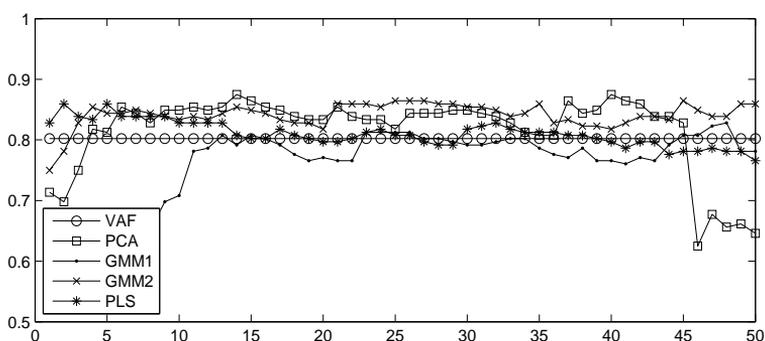


Figura 9.28: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.30: Maquinas de Vectores Soporte (RBF)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación leave-one-out:					
Precisión máxima	50.52 %	85.94 %	84.90 %	89.58 %	85.42 %
Precisión media	50.52 %	66.39 %	79.79 %	86.87 %	70.25 %
Desv. precisión	-	14.59	5.72	2.73	7.74
Sensibilidad máxima	100.00 %	85.57 %	86.60 %	89.69 %	89.69 %
Especificidad máxima	0.00 %	86.32 %	83.16 %	89.47 %	81.05 %
Validación 5-fold:					
Precisión máxima	41.67 %	85.42 %	84.90 %	90.62 %	86.46 %
Precisión media	41.67 %	64.76 %	79.61 %	86.91 %	68.31 %
Desv. precisión	-	14.44	6.06	2.63	8.75
Sensibilidad máxima	51.55 %	87.63 %	84.54 %	90.72 %	90.72 %
Especificidad máxima	31.58 %	83.16 %	85.26 %	90.53 %	82.11 %
Validación 10-fold:					
Precisión máxima	42.71 %	86.46 %	84.90 %	89.58 %	85.94 %
Precisión media	42.71 %	65.26 %	79.68 %	86.70 %	69.53 %
Desv. precisión	-	15.12	5.56	2.71	7.84
Sensibilidad máxima	52.58 %	87.63 %	85.57 %	89.69 %	89.69 %
Especificidad máxima	32.63 %	85.26 %	84.21 %	89.47 %	82.11 %

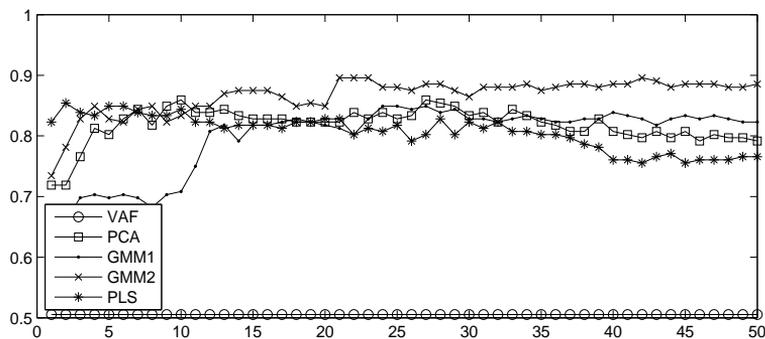


Figura 9.29: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Análisis de los resultados

Tanto la técnica con la que se han obtenido las imágenes como la composición de la base de datos son dos factores a tener en cuenta a la hora de analizar los resultados. En este caso, las imágenes son tomografías PET de controles y pacientes de AD ya que, las correspondientes a pacientes MCI no han sido consideradas para este experimento. Cabría esperar, por tanto, tasas de acierto similares a las de la base de datos CDPC. Sin embargo, hay que tener en cuenta que las imágenes etiquetadas como AD en esta base de datos, corresponden a pacientes en una fase temprana de la enfermedad, donde los efectos en la actividad cerebral son pequeños aún. Además los criterios de etiquetado en ADNI son distintos a los de las dos bases de datos anteriores. Mientras que las bases de datos HVN y CDPC han sido etiquetadas a partir de la exploración visual de las imágenes por parte de médicos experimentados, las imágenes de la base de datos ADNI se etiquetaron a partir de los resultados de varios test neuropsicológicos y sin tener en cuenta las imágenes funcionales. Dado que el diagnóstico de los sistemas CAD está basado en las imágenes, en el caso de las imágenes de HVN y CDPC, éste se ajusta mejor al diagnóstico médico (etiquetas).

Las tasas de acierto obtenidas en este experimento son ligeramente inferiores a las obtenidas con la base de datos HVN a pesar de la mayor resolución de las imágenes PET usadas en este caso. Esto se debe a que el diagnóstico basado en tests neuropsicológicos (etiquetado en ADNI) es más difícil de ajustar con un sistema CAD que el diagnóstico basado en la exploración visual de las imágenes (etiquetado en HVN).

Nuevamente el segundo enfoque del método basado en el modelo de mezcla de Gaussianas (GMM_2) arroja los mejores resultados alcanzando una precisión máxima de 90.62 % cuando se usa un clasificador Random Forests. El clasificador FLD obtiene, en este experimento, unos resultados mediocres que contrastan con los buenos resultados obtenidos con la base de datos CDPD. Y es que el escenario es ahora muy distinto: las clases no son tan fácilmente separables y un clasificador más complejo como SVM ó Random Forests puede ser ajustado satisfactoriamente puesto que el tamaño de la base de datos permite formar conjuntos de entrenamiento relativamente grandes.

En este experimento, las diferencias en las tasas de acierto de los métodos propuestos en esta Tesis Doctoral y los enfoques anteriores son más significativas. Esto pone de manifiesto la capacidad de los métodos basados en GMM y en PLS para resumir la información relevante de las imágenes funcionales en un conjunto reducido de características.

9.5.2. Controles y MCI frente a MCI converters y Alzheimer

Para este experimento han sido usadas todas las imágenes de la base de datos ADNI agrupadas en dos clases: controles más MCI y MCI converters más enfermos de AD, con 285 y 118 imágenes respectivamente. Debido al alto número de imágenes, la validación leave-one-out es demasiado costosa (desde el punto de vista del tiempo de computo) por lo que en estos experimentos se ha usado sólo la validación 10-fold y 5-fold.

Tabla 9.31: Precisión obtenida por los diferentes sistemas CAD usando todas las imágenes de la base de datos ADNI separadas en dos clases (controles + MCI y MCI converters + enfermos AD). Se muestra el valor máximo alcanzado por cada sistema, independientemente del número de características.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	56.08 %	66.00 %	60.30 %	67.00 %	59.80 %
FLD	-	63.28 %	60.55 %	62.03 %	62.78 %
NB	75.68 %	72.21 %	74.94 %	78.41 %	76.18 %
LDA	-	76.43 %	74.94 %	77.17 %	75.68 %
KNN	72.21 %	75.19 %	74.94 %	75.93 %	74.94 %
DT	61.79 %	71.71 %	70.72 %	72.95 %	73.95 %
RF	76.92 %	75.68 %	78.91 %	78.66 %	77.17 %
SVM (linear)	69.98 %	76.67 %	75.68 %	79.40 %	76.67 %
SVM (RBF)	70.72 %	77.17 %	77.67 %	79.40 %	76.67 %

En los siguientes apartados se muestra una tabla que recoge la precisión, sensibilidad y especificidad para cada uno de los clasificadores usados.

Tabla 9.32: Clasificador de la media más cercana

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	50.37 %	64.02 %	58.56 %	69.48 %	60.55 %
Precisión media	50.37 %	54.29 %	50.88 %	52.99 %	51.80 %
Desv. precisión	-	3.91	3.90	7.98	3.42
Sensibilidad máxima	53.68 %	67.72 %	61.40 %	69.82 %	67.37 %
Especificidad máxima	42.37 %	55.08 %	51.69 %	68.64 %	44.07 %
Validación 10-fold:					
Precisión máxima	56.08 %	66.00 %	60.30 %	67.00 %	59.80 %
Precisión media	56.08 %	56.32 %	52.77 %	57.99 %	51.82 %
Desv. precisión	-	3.29	3.69	4.77	3.30
Sensibilidad máxima	58.60 %	67.02 %	60.70 %	68.42 %	64.91 %
Especificidad máxima	50.00 %	63.56 %	59.32 %	63.56 %	47.46 %

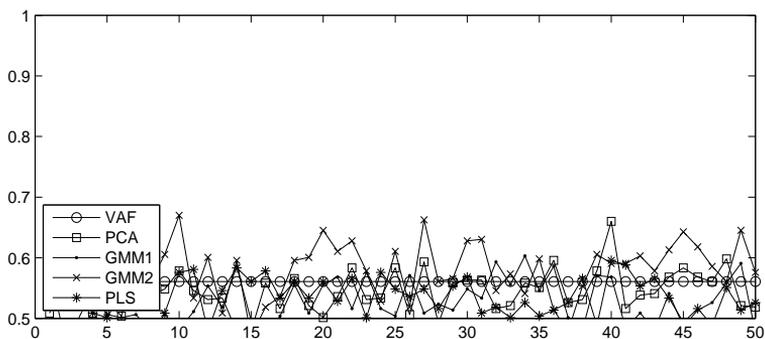


Figura 9.30: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.33: Clasificador lineal de Fisher

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	-	62.78 %	62.78 %	59.06 %	62.03 %
Precisión media	-	56.18 %	51.48 %	52.56 %	55.05 %
Desv. precisión	-	3.07	3.37	3.50	2.86
Sensibilidad máxima	-	72.98 %	65.61 %	64.91 %	70.53 %
Especificidad máxima	-	38.14 %	55.93 %	44.92 %	41.53 %
Validación 10-fold:					
Precisión máxima	-	63.28 %	60.55 %	62.03 %	62.78 %
Precisión media	-	56.41 %	52.87 %	54.19 %	55.24 %
Desv. precisión	-	2.73	2.88	2.92	2.53
Sensibilidad máxima	-	76.14 %	62.46 %	62.46 %	61.40 %
Especificidad máxima	-	32.20 %	55.93 %	61.02 %	66.10 %

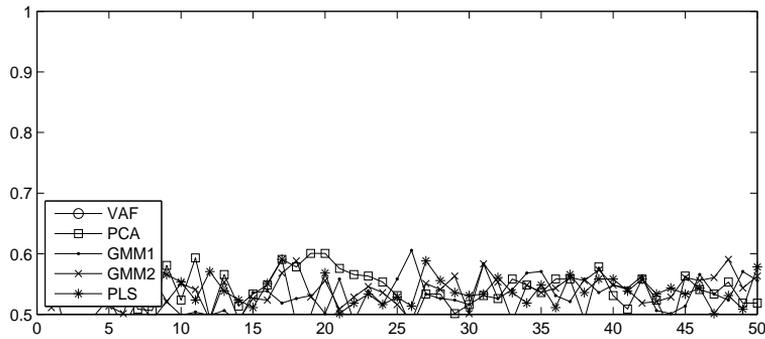


Figura 9.31: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.34: Clasificador bayesiano ingenuo

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	74.69 %	71.96 %	74.69 %	78.91 %	75.19 %
Precisión media	74.69 %	67.35 %	71.04 %	76.06 %	65.45 %
Desv. precisión	-	1.68	3.62	1.39	5.47
Sensibilidad máxima	83.51 %	77.89 %	82.81 %	85.26 %	82.11 %
Especificidad máxima	53.39 %	57.63 %	55.08 %	63.56 %	58.47 %
Validación 10-fold:					
Precisión máxima	75.68 %	72.21 %	74.94 %	78.41 %	76.18 %
Precisión media	75.68 %	67.10 %	71.15 %	76.07 %	65.12 %
Desv. precisión	-	1.59	3.47	1.32	5.60
Sensibilidad máxima	83.51 %	79.65 %	83.51 %	84.91 %	83.51 %
Especificidad máxima	56.78 %	54.24 %	54.24 %	62.71 %	58.47 %

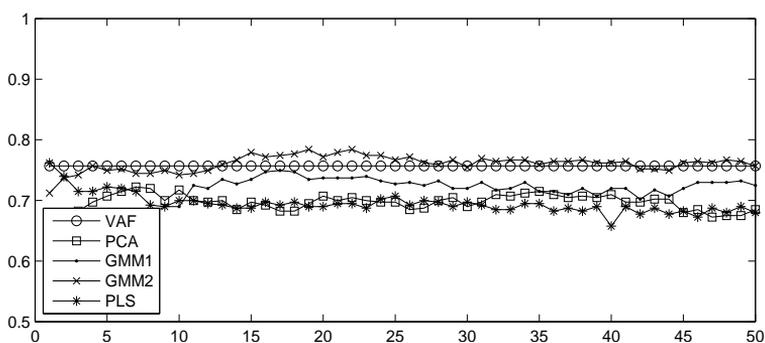


Figura 9.32: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.35: Análisis Discriminante Lineal

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	-	76.43 %	75.19 %	76.92 %	75.68 %
Precisión media	-	66.88 %	69.63 %	74.30 %	69.96 %
Desv. precisión	-	5.80	5.40	1.85	2.86
Sensibilidad máxima	-	78.60 %	79.30 %	80.35 %	79.30 %
Especificidad máxima	-	71.19 %	65.25 %	68.64 %	66.95 %
Validación 10-fold:					
Precisión máxima	-	76.43 %	74.94 %	77.17 %	75.68 %
Precisión media	-	67.20 %	69.78 %	74.55 %	71.09 %
Desv. precisión	-	5.80	5.45	1.42	2.25
Sensibilidad máxima	-	80.70 %	81.05 %	80.35 %	78.25 %
Especificidad máxima	-	66.10 %	60.17 %	69.49 %	69.49 %

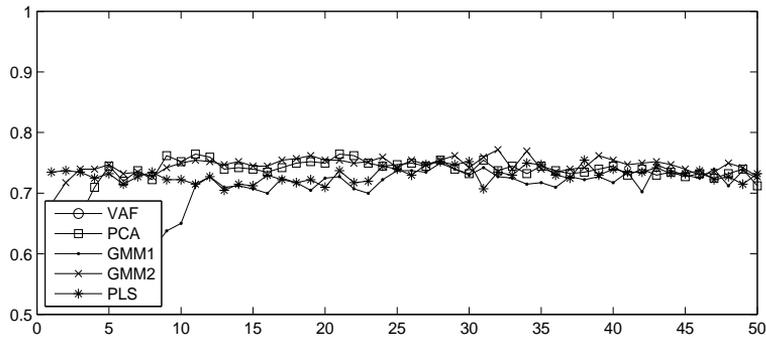


Figura 9.33: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.36: Clasificador de los k vecinos más cercanos

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	72.46 %	75.93 %	74.69 %	76.67 %	73.95 %
Precisión media	72.46 %	72.64 %	71.81 %	73.86 %	66.48 %
Desv. precisión	-	1.24	2.47	1.64	3.03
Sensibilidad máxima	92.98 %	94.04 %	91.58 %	86.67 %	89.82 %
Especificidad máxima	22.88 %	32.20 %	33.90 %	52.54 %	35.59 %
Validación 10-fold:					
Precisión máxima	72.21 %	75.19 %	74.94 %	75.93 %	74.94 %
Precisión media	72.21 %	72.93 %	72.01 %	73.61 %	66.37 %
Desv. precisión	-	1.08	2.50	1.40	3.06
Sensibilidad máxima	93.33 %	90.88 %	90.88 %	86.67 %	89.47 %
Especificidad máxima	21.19 %	37.29 %	36.44 %	50.00 %	39.83 %

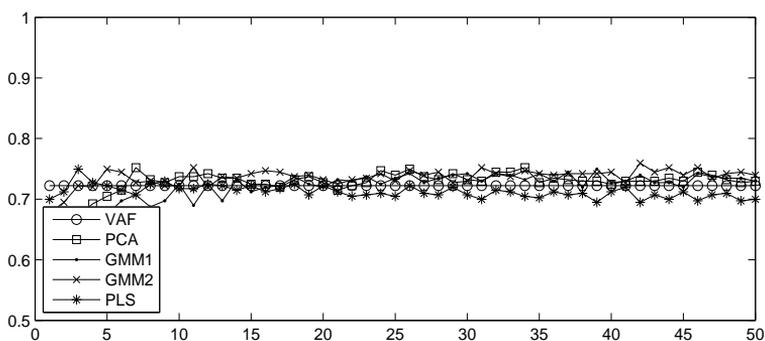


Figura 9.34: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.37: Árboles de Decisión

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	66.25 %	70.22 %	70.97 %	74.19 %	71.96 %
Precisión media	66.25 %	63.30 %	66.30 %	68.82 %	66.20 %
Desv. precisión	-	2.50	3.03	2.00	2.21
Sensibilidad máxima	74.74 %	77.89 %	78.60 %	81.75 %	81.40 %
Especificidad máxima	45.76 %	51.69 %	52.54 %	55.93 %	49.15 %
Validación 10-fold:					
Precisión máxima	61.79 %	71.71 %	70.72 %	72.95 %	73.95 %
Precisión media	61.79 %	63.65 %	66.69 %	69.30 %	66.15 %
Desv. precisión	-	2.44	2.64	1.68	2.21
Sensibilidad máxima	70.53 %	81.05 %	81.40 %	81.75 %	81.05 %
Especificidad máxima	40.68 %	49.15 %	44.92 %	51.69 %	56.78 %

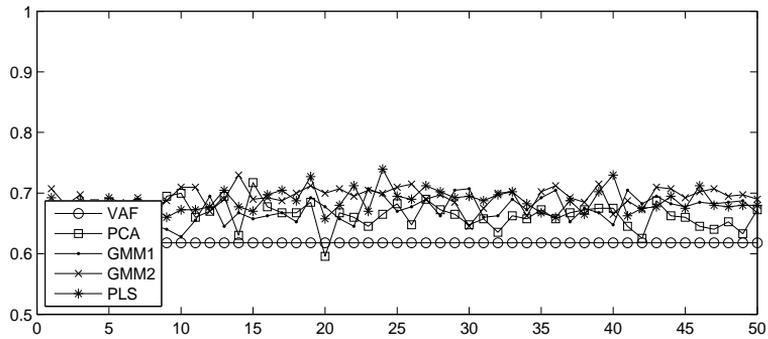


Figura 9.35: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.38: Random Forests

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	73.20 %	75.93 %	77.67 %	79.40 %	77.67 %
Precisión media	73.20 %	71.48 %	73.09 %	75.73 %	72.79 %
Desv. precisión	-	1.36	4.26	2.42	1.35
Sensibilidad máxima	89.47 %	94.04 %	91.23 %	89.82 %	97.54 %
Especificidad máxima	33.90 %	32.20 %	44.92 %	54.24 %	29.66 %
Validación 10-fold:					
Precisión máxima	76.92 %	75.68 %	78.91 %	78.66 %	77.17 %
Precisión media	76.92 %	71.41 %	73.34 %	75.29 %	72.92 %
Desv. precisión	-	1.32	4.09	2.29	1.31
Sensibilidad máxima	92.28 %	92.63 %	91.58 %	89.47 %	95.44 %
Especificidad máxima	39.83 %	34.75 %	48.31 %	52.54 %	33.05 %

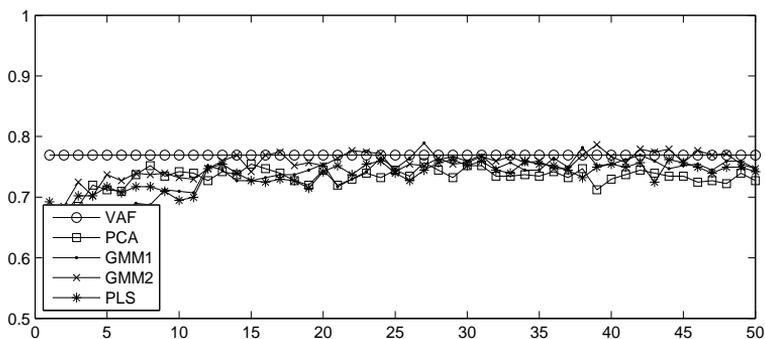


Figura 9.36: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.39: Maquinas de Vectores Soporte (lineal)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	68.24 %	77.42 %	77.17 %	78.91 %	76.92 %
Precisión media	68.24 %	70.30 %	73.83 %	76.36 %	73.68 %
Desv. precisión	-	5.39	2.00	1.37	1.88
Sensibilidad máxima	78.60 %	87.72 %	91.58 %	90.88 %	88.42 %
Especificidad máxima	43.22 %	52.54 %	42.37 %	50.00 %	49.15 %
Validación 10-fold:					
Precisión máxima	69.98 %	76.67 %	75.68 %	79.40 %	76.67 %
Precisión media	69.98 %	72.09 %	73.49 %	77.06 %	73.99 %
Desv. precisión	-	4.61	1.58	1.10	1.75
Sensibilidad máxima	81.05 %	86.32 %	90.53 %	90.88 %	89.82 %
Especificidad máxima	43.22 %	53.39 %	39.83 %	51.69 %	44.92 %

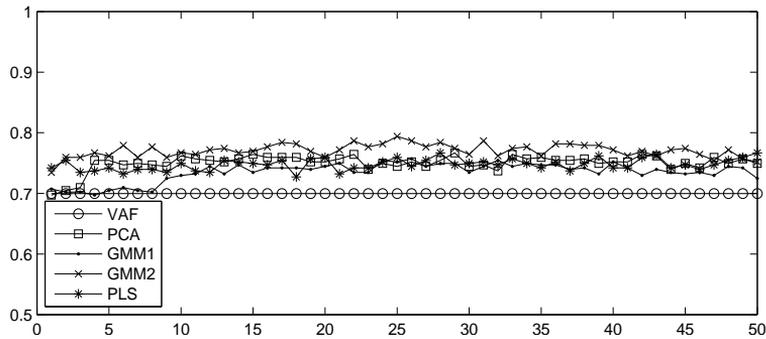


Figura 9.37: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

Tabla 9.40: Maquinas de Vectores Soporte (RBF)

	VAF	PCA	GMM ₁	GMM ₂	PLS
Validación 5-fold:					
Precisión máxima	70.72 %	78.16 %	77.67 %	78.41 %	76.18 %
Precisión media	70.72 %	71.40 %	74.52 %	76.99 %	71.29 %
Desv. precisión	-	1.67	1.95	0.87	1.35
Sensibilidad máxima	100.00 %	94.39 %	90.88 %	90.88 %	92.98 %
Especificidad máxima	0.00 %	38.98 %	45.76 %	48.31 %	35.59 %
Validación 10-fold:					
Precisión máxima	70.72 %	77.17 %	77.67 %	79.40 %	76.67 %
Precisión media	70.72 %	71.42 %	74.85 %	77.10 %	71.33 %
Desv. precisión	-	1.65	1.94	1.03	1.42
Sensibilidad máxima	100.00 %	91.23 %	90.88 %	90.88 %	92.28 %
Especificidad máxima	0.00 %	43.22 %	45.76 %	51.69 %	38.98 %

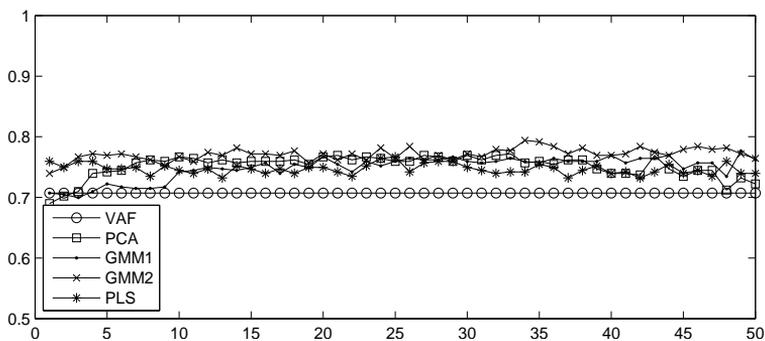


Figura 9.38: Comparativa de la precisión de los diferentes sistemas CAD en función del número de componentes

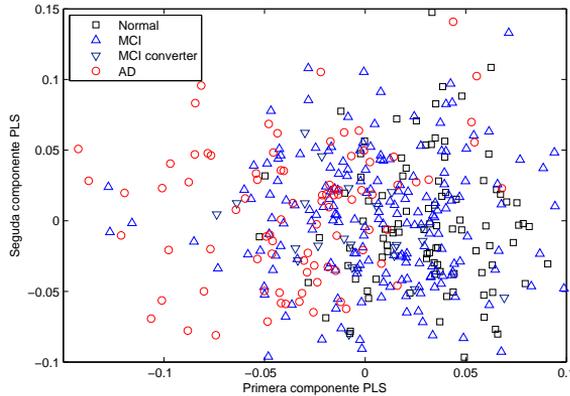


Figura 9.39: Representación de las 403 imágenes de la base de datos ADNI mediante sus las dos primeras componentes PLS. Obsérvese como las imágenes MCI están muy mezcladas con los controles y es difícil distinguir entre MCI y MCI converters.

Análisis de los resultados

Distinguir entre MCI y MCI converter es una tarea muy compleja como muestran las bajas tasas de acierto obtenidas por los diferentes métodos empleados. En la figura 9.39 se representan las dos primeras componentes PLS de todas las imágenes de la base de datos ADNI. Esta figura pone de manifiesto la variabilidad del patrón MCI, lo que hace muy difícil separar satisfactoriamente las imágenes de este tipo.

Otro de los problemas que influye en la baja precisión de los sistemas CAD es el reducido número de ejemplos de tipo MCI converters. A pesar de contar con muchas imágenes, la base de datos está claramente descompensada en este sentido. Aún así se obtienen tasas de acierto que rozan el 80%. Destaca el hecho de que sea un clasificador no lineal como SVM-RBF el que consigue los mejores resultados. Incluso con VAF, donde el elevado número de características suele perjudicar notablemente a los clasificadores no lineales, en este caso se obtienen resultados aceptables: una precisión del 70.72%, tan solo 6 puntos porcentuales por debajo del mejor clasificador para este método. Este comportamiento se debe fundamentalmente al mayor tamaño de los conjuntos de entrenamiento usados para la validación que permite un mejor ajuste de la función de clasificación.

9.6. Discusión

Las tablas 9.1, 9.11, 9.21 y 9.31 muestran como los dos métodos propuestos en este trabajo mejoran, en la mayoría de los casos, la precisión obtenida por enfoques anteriores. Estos resultados demuestran que la selección de características basada en la selección automática de ROIs y en la técnica de mínimos cuadros parciales permite reducir la dimensionalidad del espacio de características conservando la información más relevante de las imágenes funcionales. En este sentido, conviene destacar las diferencias existentes entre las dos variantes del método basado en GMM, a saber, la primera aproximación que usa una media de controles para extraer las ROIs (GMM_1) y su posterior refinamiento que emplea una imagen con las diferencias entre controles e imágenes patológicas (GMM_2). Mientras que el enfoque denominado GMM_1 consigue tasas de acierto similares a las conseguidas por el método basado en PCA, GMM_2 supera significativamente esos valores. Si comparamos los enfoques GMM_2 y PLS, vemos que GMM_2 se comporta mejor con las bases de datos CDPC y ADNI formadas por imágenes PET mientras que PLS alcanza mejores resultados con imágenes SPECT como las de la base de datos HVN. Además, el método basado en PLS tiene menores tiempos de ejecución.

Los experimentos realizados nos permiten además encontrar los clasificadores más apropiados para cada método. En general SVM y RF son los clasificadores que mejor se comportan, si bien, FLD obtiene los mejores resultados con la base de datos CDPC como se ha visto en la sección 9.4.1.

En la construcción de un sistema CAD, no sólo es importante la precisión que se alcance, sino también la sensibilidad y especificidad que consiga. Las curvas ROC representadas en la figura 9.40 muestran como los métodos propuestos alcanzan el equilibrio necesario entre estas dos medidas y lo comparan con el conseguido por VAF y PCA. Para generarlas se ha elegido el clasificador con el que mejores resultados conseguía cada método, esto es, SVM-lineal para todos los casos salvo para el modelo basado en GMM, donde el clasificador SVM-RBF ha producido mejores resultados. Los diferentes resultados de cada método se obtienen al variar el número de Gaussianas/componentes usadas en la extracción de características, es decir, el tamaño de los vectores de características. Obsérvese como para la base de datos CDPC la especificidad es más elevada que la sensibilidad debido a lo descompensada que está esta base de datos (sólo 13 controles y 47 imágenes patológicas). En cambio para el segundo experimento de la base de datos ADNI sucede lo contrario, la sensibilidad es mucho más elevada que la especificidad, debido al mismo motivo (en este caso hay 285 controles y sólo 118 imágenes patológicas).

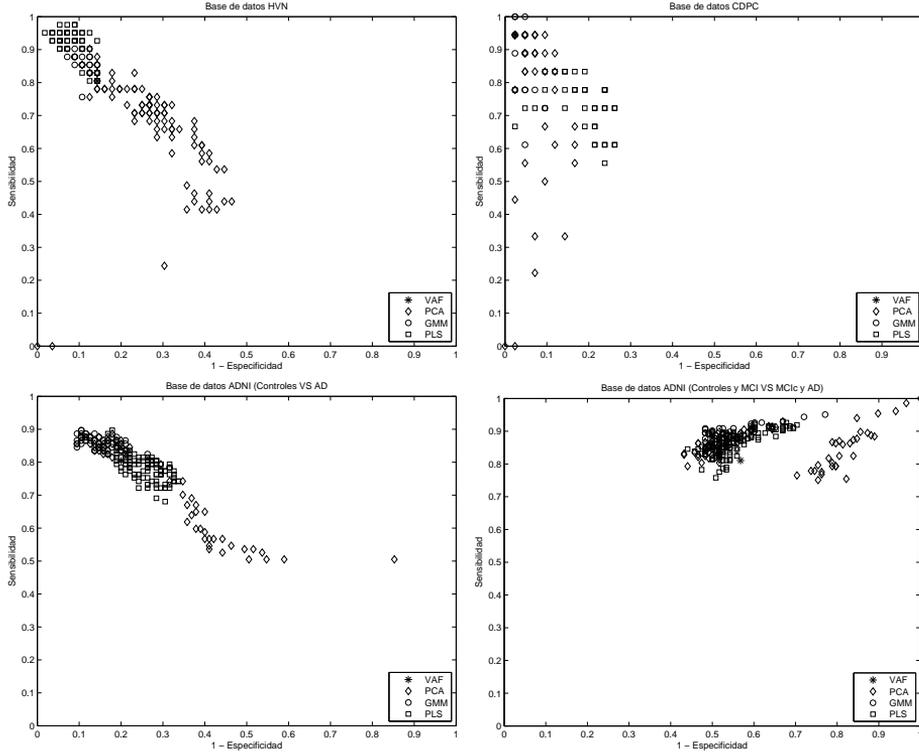


Figura 9.40: Curvas ROC de los métodos VAF, PCA, GMM_2 y PLS para las tres bases de datos usadas en este trabajo. Todos los métodos usan un clasificador SVM-lineal salvo GMM_2 que usa SVM-RBF

El uso de dos *kernel*s distintos con SVM nos permite comparar un clasificador lineal (el que usa *kernel* lineal) con otro no lineal (el que usa *kernel* RBF). En general, un clasificador lineal requiere menos muestras durante el proceso de entrenamiento para obtener resultados aceptables. Esto equivale a decir que, a igual número de muestras, un clasificador no lineal será más sensible al problema del pequeño tamaño muestral y para conseguir buenos resultados de clasificación será necesario obtener vectores de características de menor dimensión que en el caso de clasificadores lineales. Por este motivo, el método basado en GMM, que es el que realiza una reducción más drástica de la dimensionalidad (sin trunca

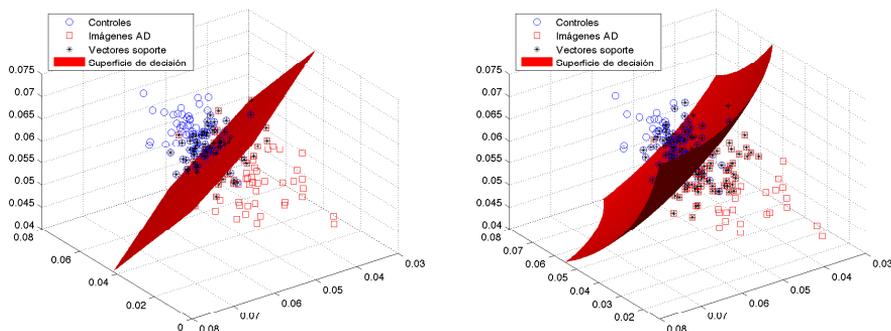


Figura 9.41: Superficie de decisión de un clasificador SVM con *kernel* lineal (izquierda) y de SVM con *kernel* RBF (derecha). Se muestran las tres primeras características de las imágenes de la base de datos ADNI usando el método GMM_2

los vectores de características), obtiene mejores resultados con un clasificador SVM-RBF que con SVM-lineal, mientras que, para el resto de métodos, la variante lineal obtiene tasas de precisión más altas. Obsérvense por ejemplo los resultados para el método VAF, que destaca por la elevada dimensión de los vectores de características. Mientras SVM-lineal consigue una tasa de acierto del 83.51 %, SVM-RBF sólo obtiene una precisión del 57.73 % para la base de datos HVN y para el resto de base de datos, las diferencias son similares. La figura 9.41 muestra gráficamente como el clasificador no lineal separa mejor los datos a partir del método GMM_2 . Para la representación se han usado las tres primeras posiciones de los vectores de características generados con el modelo GMM_2 y las imágenes de la base de datos ADNI usadas en el experimento descrito en la sección 9.5.1.

CAPÍTULO

10

HERRAMIENTA PARA EL ESTUDIO DE IMÁGENES MÉDICAS

En este capítulo se muestra el desarrollo de una herramienta de ayuda al diagnóstico (CAD) para enfermedades neurodegenerativas implementada sobre Matlab. Con ella se pretende hacer llegar a los hospitales y centros médicos los últimos avances en el diagnóstico asistido de enfermedades neurológicas tales como la enfermedad de Alzheimer. Esta herramienta representa un sistema completo de tratamiento de imágenes tomográficas ya que está pensada para recibir las imágenes tal cual son generadas por la cámara, normalizarlas tanto en intensidad como especialmente de manera semiautomática y emitir un diagnóstico en base a diferentes métodos de aprendizaje estadístico.

10.1. Introducción

Uno de los retos al que se enfrenta todo avance científico es llegar al público, es decir, pasar la barrera del laboratorio y ser implantado en la sociedad. En el caso que nos ocupa, la llegada de los sistemas CAD a los hospitales y centros médicos suele ser lenta y en ocasiones supone un importante esfuerzo de aprendizaje por parte de los médicos. Posiblemente no haya un motivo principal que justifique esta adopción tardía sino que, más bien, se trata de un cúmulo de varios factores negativos. Por un lado, los científicos e investigadores a menudo son incapaces de implementar y describir los sistemas de ayuda al diagnóstico en un lenguaje cercano al usado en medicina, lo que hace que los médicos no lleguen a comprender la utilidad de dichos sistemas. Por otro lado los médicos en muchas ocasiones se muestran reticentes a confiar en una máquina cuyo funcionamiento interno desconocen salvo cuando ésta ya está ampliamente aceptada y es usada en un elevado número de centros. A estos inconvenientes hay que sumar los problemas derivados de patentes, problemas económicos, etc.

Para salvar las dificultades arriba mencionadas y hacer llegar los avances recientes en el campo del diagnóstico asistido por computador, como son los métodos descritos en los capítulos 7 y 8, hemos desarrollado una herramienta software que permite acceder a ellos mediante una interfaz sencilla. Sus principales características son las siguientes:

- Normalización automática de las imágenes, tanto espacialmente como en intensidad.
- Varias formas de visualización que facilitan la comparación de imágenes. Además es posible aplicar diferentes mapas de color para resaltar las diferencias.
- Posibilidad de re-alimentar los algoritmos de aprendizaje estadístico con nuevos conjuntos de imágenes etiquetadas.
- Permite usar simultáneamente varios métodos de diagnóstico, es decir, se pueden usar varios métodos para obtener varias valoraciones (diagnósticos) de una misma imagen.
- Diseño modular que permite añadir fácilmente nuevos métodos que puedan desarrollarse en un futuro.

El hecho de estar implementada sobre Matlab [Attaway, 2009] convierte a esta herramienta en multiplataforma ya que existen versiones de Matlab para

Windows, MacOS y Linux, al tiempo que facilita la adición de nuevos métodos (los algoritmos se escriben en un lenguaje interpretado que no requiere compilación). Además es posible portar el código a lenguaje C de manera semi-automática en caso de que fuese necesario ejecutar la herramienta en entornos no soportados por Matlab.

10.2. Formatos de archivo

Las imágenes médicas suelen usar formatos de imagen específicos que permitan codificar imágenes tridimensionales y almacenar junto a la intensidad de cada vóxel ciertas etiquetas con información relativa a la imagen, como el nombre del paciente al que pertenece, el tipo de imagen, la cámara y el trazador usado, etc. Por este motivo, existen formatos específicos, ampliamente extendidos en la comunidad científico-médica, como DICOM o Analyze. Los formatos soportados por la herramienta que presentamos son los siguientes:

- **Mapa de bits (BMP).** Es un formato de propósito general por lo que no permite almacenar información adicional junto a la imagen y sólo es válido para imágenes en 2 dimensiones, por lo tanto, para almacenar una imagen tridimensional es necesario usar varios archivos de mapas de bits.
 - **Digital Imaging and Communications in Medicine (DICOM).** Es un estándar reconocido mundialmente para el intercambio de imágenes médicas y pensado para el manejo, almacenamiento, impresión y transmisión de estas imágenes. Incluye la definición de un formato de fichero y de un protocolo de comunicación de red. El protocolo de comunicación es un protocolo de aplicación que usa TCP/IP para la comunicación entre sistemas. Los ficheros DICOM pueden intercambiarse entre dos entidades que tengan capacidad de recibir imágenes y datos de pacientes en formato DICOM. La primera versión fue desarrollada en 1985 y actualmente es uno de los formatos más extendidos.
 - **Mayo/Analyze.** Analyze es un paquete software desarrollado por el Biomedical Imaging Resource (BIR) en la clínica Mayo para la visualización y procesamiento de imágenes biomédicas multidimensionales. Este software introdujo un nuevo formato de imagen que almacena por separado la cabecera (información relativa a la imagen) y la imagen, usando por tanto dos archivos diferentes para cada imagen. El uso de este formato está muy extendido y es usado por programas desarrollados posteriormente como SPM.
-

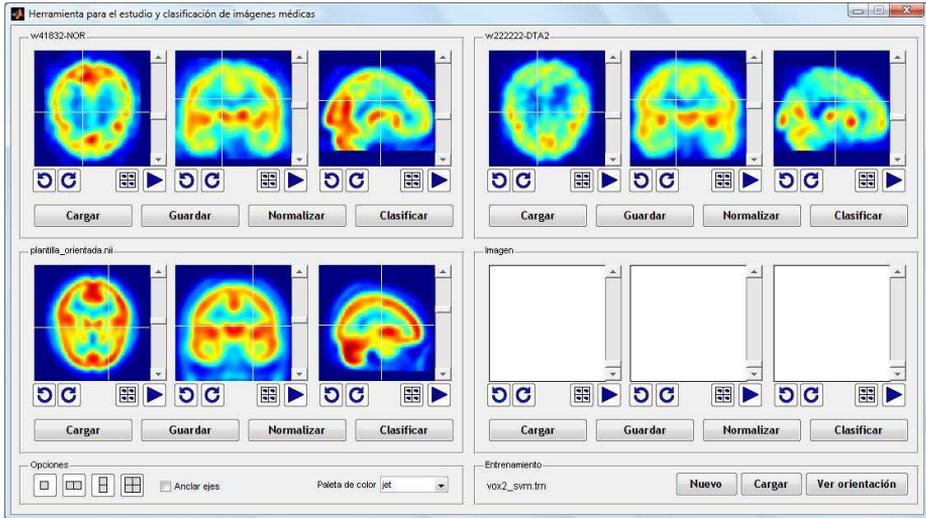


Figura 10.1: Ventana principal de la herramienta

- **NIfTI-1.** Propuesto por la iniciativa NIfTI (Neuroimaging Informatics Technology Initiative), es una extensión del formato MAYO/Analyze que añade campos de información a la cabecera y permite almacenar cabecera e imagen en un solo archivo facilitando el manejo de las imágenes. Al igual que el formato al que extiende, es soportado por SPM.

10.3. Visualización y preprocesado de imágenes

La normalización es el proceso que garantiza que dos imágenes distintas sean comparables. Como es obvio, imágenes procedentes de distintos pacientes no tienen *a priori* la misma forma ni el mismo tamaño por lo que no es posible una comparación directa entre ellas. Para garantizar que las mismas posiciones anatómicas se correspondan con las mismas posiciones en la imagen, es necesario normalizar espacialmente ambas imágenes *deformándolas* hasta hacerlas coincidir con una plantilla. Además, pequeñas variaciones en la cantidad del radiofármaco inyectado a cada paciente, así como la diferente respuesta de cada persona ante el radiofármaco, pueden hacer que la intensidad de algunas imágenes varíe entre el 0 y el 50 %, mientras que en otras lo haga entre el 0 y el 90 %.

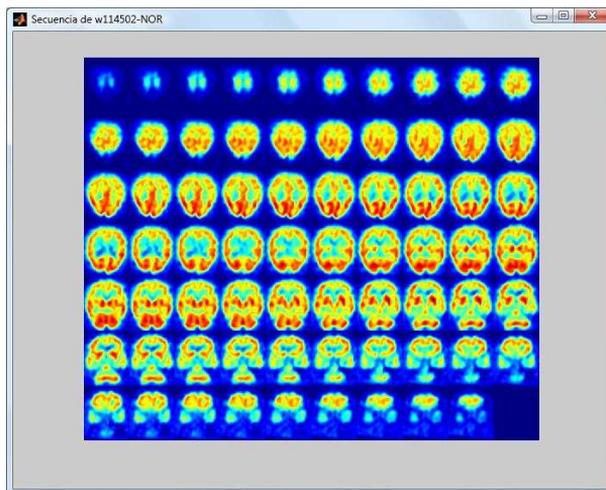


Figura 10.2: Secuencia de una imagen SPECT en la dimensión axial

La normalización en intensidad elimina estas diferencias igualando el nivel de los vóxeles de mayor intensidad y reescalando los restantes.

La figura 10.1 muestra la ventana principal de la herramienta. Su interfaz simple y personalizable permite comparar visualmente hasta cuatro imágenes mostrando tres cortes (uno por dimensión) de cada una de ellas. Las opciones disponibles para cada imagen son las siguientes:

- **Normalizar.** Lleva a cabo la normalización espacial de la imagen. Durante el proceso se solicita que se indique una imagen que se usará como plantilla para normalizar.
- **Rotar.** Junto a cada corte de la imagen aparecen dos botones para girar la imagen a izquierda o derecha.
- **Secuenciar.** Permite obtener una visión unidimensional de la imagen en la que se muestran correlativos una serie de cortes de la imagen tridimensional, en una dimensión determinada. Puede verse un ejemplo en la figura 10.2. Este tipo de representación es a menudo usada por los médicos para la exploración manual.
- **Animación.** Muestra sucesivamente todos los cortes en la dimensión seleccionada. Es un efecto similar a una secuencia de vídeo en la que se

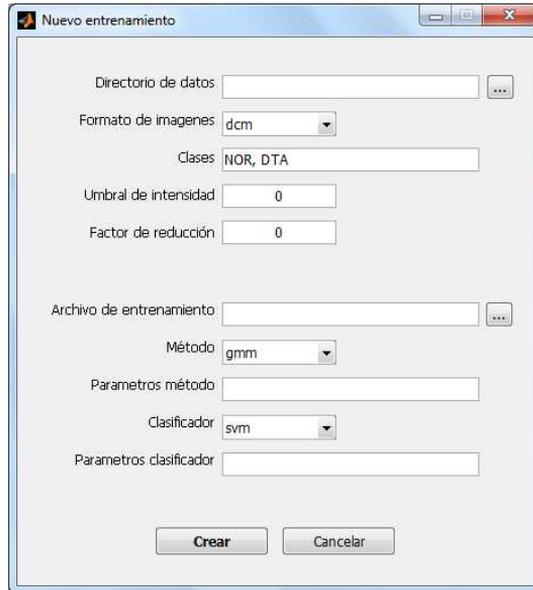


Figura 10.3: Cuadro de diálogo para construir un nuevo entrenamiento

recorre el cerebro en una dimensión.

- **Clasificar.** Permite obtener un diagnóstico automático a partir de la imagen. Esta función requiere que previamente se haya indicado un entrenamiento.

10.4. Entrenamiento y clasificación

En este contexto, un entrenamiento es un archivo en el que se almacena una determinada función de clasificación (véase capítulo 5) que se ha ajustado siguiendo un proceso de aprendizaje supervisado mediante las imágenes etiquetadas que se indicaron en el momento de su creación. La finalidad de estos entrenamientos es poder conseguir un diagnóstico en tiempo real para una imagen determinada, sin tener que repetir la fase de ajuste del clasificador.

Los entrenamientos se construyen desde un diálogo específico (figura 10.3) en el que se debe indicar el método de extracción de características, el tipo de

clasificador y los ejemplos etiquetados para ajustar la función de clasificación. El proceso de construcción de un entrenamiento sigue los siguientes pasos:

- Leer y procesar las imágenes usadas como ejemplos.
- Determinar las etiquetas asociadas a los ejemplos.
- Extraer los vectores de características correspondientes a los ejemplos mediante uno de los métodos de los que está dotada la herramienta.
- Calcular los parámetros de la función de clasificación de acuerdo al clasificador elegido.
- Guardar los parámetros calculados en un archivo para su posterior uso.

Este diseño permite disponer de varios entrenamientos diferentes (con distintos métodos de extracción de características, distintos clasificadores, etc.) y usar el más adecuado en cada momento, dependiendo del tipo de imagen sobre la que queramos obtener un diagnóstico.

Debido a la complejidad de algunos métodos de extracción de características, la creación de un entrenamiento es un proceso computacionalmente costoso que puede requerir desde unos pocos minutos hasta días completos.

CAPÍTULO

11

CONCLUSIONES Y TRABAJO FUTURO

En este capítulo se recogen las conclusiones de la Tesis Doctoral resaltando las aportaciones científicas realizadas. Dichas aportaciones se dividen en dos grupos. Por un lado, el desarrollo de nuevos algoritmos que mejoran la precisión de los sistemas CAD para enfermedades neurodegenerativas tales como la enfermedad de Alzheimer y, por otro lado, la implementación de una herramienta software que acerque a los hospitales los últimos logros obtenidos en materia de sistemas CAD. Asimismo, en este capítulo se realizan una serie de propuestas para continuar este trabajo y abrir nuevas líneas de investigación que mejoren los métodos aquí propuestos.

11.1. Conclusiones

Las principales aportaciones de esta Tesis Doctoral se dividen en dos vertientes. Por un lado, el desarrollo de nuevos métodos para el análisis de imágenes funcionales tomográficas, que permiten mejorar la precisión de los sistemas CAD usados para el diagnóstico de enfermedades neurodegenerativas y, por otro lado, la implementación de una herramienta software que acerque los últimos avances en sistemas CAD, como los métodos expuestos en este trabajo, a los centros médicos.

Más concretamente se han desarrollado dos métodos para el análisis de imágenes funcionales. En el capítulo 7 se ha presentado un método que hace uso del Modelo de Mezcla de Gaussianas para agrupar o parcelar los vóxeles de las imágenes, definiendo de esta forma, regiones de interés. Las regiones de interés se definen sobre una imagen que denominamos *imagen modelo* y que puede ser definida de varias formas. Inicialmente construimos esta imagen como una media de controles, lo que nos lleva a definir como regiones de interés aquellas zonas del cerebro que tienen mayor actividad en personas sanas. Posteriormente, en aras de mejorar los resultados del diagnóstico automático, decidimos definir las regiones de interés como aquellas zonas en las que la actividad media de las personas sanas y de las personas enfermas difiere. De esta forma conseguimos que el análisis se centre sólo en las zonas que son útiles para distinguir controles de imágenes patológicas. Para ello, construimos la *imagen modelo* como la resta vóxel a vóxel entre la imagen media de los controles y la imagen media de las imágenes patológicas. Una vez construida esta imagen se aplica un algoritmo de mezcla de Gaussianas, que modela las regiones de interés como una suma de k Gaussianas de forma que el centro, la varianza y el peso de las Gaussianas representan respectivamente la posición, la forma y la importancia de las regiones de interés. A partir de estos valores construimos, para cada imagen, un vector de características reducido y que contiene toda la información relevante de la imagen. Este vector contiene un elemento por cada una de las k Gaussianas y se calcula como la activación de la imagen completa para esa Gaussiana.

El segundo de los métodos de análisis de imágenes funcionales presentados aparece descrito en el capítulo 8 y está basado en la técnica de Mínimos Cuadrados Parciales. Esta técnica, similar a PCA, descompone dos conjuntos de variables como el producto de dos matrices denominadas *scores* y *loadings* siguiendo un criterio de maximización de la covarianza. En nuestro caso, los dos conjuntos de variables son los formados por las imágenes funcionales y por las etiquetas de dichas imágenes. Tras la descomposición de estos conjuntos, la matriz de *scores* correspondiente al conjunto de las imágenes contiene proyecciones

de los vóxeles que, a diferencia de PCA, han sido obtenidas teniendo en cuenta las etiquetas de las imágenes. Estas proyecciones conocidas como componentes PLS son usadas como vectores de características.

El análisis de las imágenes funcionales usando los dos métodos propuestos, tiene como fin extraer vectores de características reducidos que, usados para entrenar un clasificador estadístico, nos permitan mejorar la precisión de los sistemas CAD para enfermedades neurodegenerativas. Por este motivo, estos métodos han sido validados mediante cuatro experimentos distintos en los que se hace uso de imágenes SPECT y PET de tres bases de datos para el diagnóstico de la enfermedad de Alzheimer. Concretamente se usan 97 imágenes SPECT proporcionadas por el Hospital Virgen de las Nieves (Granada), 60 imágenes PET procedentes del centro de diagnóstico PET Cartuja (Sevilla) y 403 imágenes PET seleccionadas de la base de datos de la iniciativa ADNI.

Para la construcción de los sistemas CAD se han usado 8 clasificadores estadísticos de diferentes tipos, a saber, clasificador de la media más cercana, clasificador lineal de Fisher, clasificador bayesiano ingenuo, análisis discriminante lineal, clasificador de los k vecinos más cercanos, árboles de decisión, random forests y máquinas de vectores soporte. Las tasas de acierto de los sistemas CAD han sido estimadas mediante validación leave-one-out y k -fold y los resultados obtenidos son comparados con enfoques anteriores como VAF y PCA. Estos resultados se detallan en el capítulo 9 y se pueden resumir de la siguiente forma:

- El método basado en PLS consigue una precisión en el diagnóstico de AD usando imágenes SPECT cercana al 97%, lo que supone una mejora de 8 puntos porcentuales sobre la precisión obtenida con PCA y 13 puntos porcentuales sobre la precisión obtenida con VAF. Este método se muestra igualmente efectivo en la clasificación de imágenes PET, sin embargo en este caso las diferencias con enfoques anteriores no son tan abultadas.
- El método basado en GMM se muestra muy estable alcanzando resultados significativamente mejores que los enfoques anteriores con todos los clasificadores empleados. Destaca sobre todo en la clasificación de las imágenes de la base de datos ADNI donde el criterio de etiquetado y la composición de la base de datos hacen de la clasificación una tarea especialmente compleja. Además este método demuestra un comportamiento notable con un clasificador no lineal como SVM con *kernel* RBF.

Finalmente, en el capítulo 10 se ha mostrado una herramienta software destinada a llevar a los centros médicos, los sistemas CAD desarrollados en este

trabajo. La implementación se ha realizado sobre Matlab, lo que convierte esta herramienta en multiplataforma en tanto en cuanto, puede ejecutarse sobre diferentes sistemas operativos. Además, su diseño modular permite añadir fácilmente nuevos métodos en el futuro.

11.2. Trabajo futuro

La investigación realizada en este trabajo abre las puertas a nuevas líneas de investigación que completen y mejoren el mismo. Estas nuevas líneas se pueden agrupar en 4 apartados:

- **Fusión de imágenes.** Una de las desventajas de PET sobre SPECT es su mayor coste, lo que hace que, en la práctica, la técnica SPECT esté mucho más extendida a pesar de que las imágenes que genera tienen menor resolución que las PET. Por otra parte, las imágenes de resonancia magnética tienen una alta resolución pero menor capacidad para reflejar los procesos metabólicos. La fusión de imágenes SPECT y MRI puede convertirse en una alternativa a PET, económica y capaz de producir imágenes de alta resolución.
 - **Mejoras en normalización.** Las técnicas de análisis de imagen funcional como las mostradas en este trabajo tienen una gran dependencia de la normalización que se ha realizado sobre las imágenes en la etapa de preprocesado. Por este motivo el desarrollo de nuevas técnicas de normalización más precisas redundará en mejoras en los sistemas de clasificación.
 - **Fusión de métodos.** El método basado en Mezcla de Gaussianas selecciona el número de Gaussianas manteniendo un equilibrio entre capacidad de reconstrucción de la imagen original (requiere un número elevado de Gaussianas) y tamaño reducido de los vectores de características (requiere un número pequeño de Gaussianas). Una posible alternativa consiste en generar un modelo con un número elevado de Gaussianas, consiguiendo una alta capacidad de reconstrucción y, posteriormente, seleccionar sólo algunas Gaussianas mediante algún método alternativo como PCA o PLS.
 - **Mejoras en la herramienta.** Sin duda, la herramienta informática desarrollada tiene un amplio margen de mejora, dotándola de más métodos y adaptándola a los requerimientos de los usuarios, una vez haya sido implementada en hospitales y se reciba el *feedback* de los médicos.
-

Part IV

Summary in English

Abstract

In this Thesis two new methods for analyzing functional images are proposed for computer aided diagnosis of neurological diseases, such as Alzheimer's disease (AD).

The first method proposed uses the Gaussian Mixture Model (GMM) to cluster or parcel voxels that are similar in a certain way. First, an image which contains regions of interest (ROIs) is computed. This image can be defined in several ways: as the average of normal images or as the difference between normal and pathological images. Then a GMM algorithm is applied to model this image with a mixture of Gaussian distributions. Once Gaussians have been defined, they are used to compute a feature vector for each functional image. This vector consists of one element for each Gaussian of the mixture, which is calculated as the activation of the entire image for that Gaussian.

The second method proposed is based on the Partial Least Squares approach. This technique transforms two sets of variables into the product of two matrices called *scores* and *loadings* according to a criterion of covariance maximization. In this case, a set consists of the functional images and the other one consists of the labels of these images. After transformation, the *score* matrix for the set of images consists of projections of voxels that have been obtained by taking into account the labels of the images. These projections, named PLS components, are used as feature vectors.

These methods have been combined with several supervised classification techniques and the accuracy rates of the resulting CAD systems have been estimated by means of 3 different databases of SPECT and PET images for AD. The results obtained outperform previous approaches such as the ones based on Voxel-As-Features and Principal Component Analysis.

Finally, we have implemented a software tool that allows to manage and analyze functional images using the methods presented in this Thesis and other previous approaches. This tool has an intuitive graphical interface and is easy to use.

CHAPTER

12

INTRODUCTION

Dementia, one of the most severe and frequent neurodegenerative disorders in the elderly population, has important and dramatic health as well as socio-economic implications [Ritchie & Lovestone, 2002]. Furthermore, the incidence and prevalence of these diseases is increasing due to the aging population. In recent years, the use of Computer-Aided Diagnosis systems based on tomographic images of the brain has allowed significant improvements in the diagnosis of several dementias. However, processing tomographic images entails some problems which should be addressed.

12.1. Computer-Aided Diagnosis systems

Several authors [Thompson & Apostolova, 2007; Whitwell *et al.*, 2007; Johnson *et al.*, 2006] suggest that pathological manifestations of neurodegenerative diseases begin many years before the patient becomes symptomatic. In order to anticipate the diagnosis and thus to improve treatments, medical imaging techniques based on nuclear medicine are a significant advance. These techniques obtain three-dimensional maps of brain activity which allow to perform a diagnosis based on hipoperfusion or hipometabolism pattern of the brain. Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) are tomographic imaging modalities largely used for the study of the regional Cerebral Blood Flow (rCBF) and glucose metabolism, respectively.

The analysis of functional images (FI) using computers can be performed at several scales. On the one hand, the most familiar scale to the neuroimaging community concerns mass univariate statistical testing, which models data at the scale of individual voxel. In the early 2000's it was demonstrated [Signorini *et al.*, 1999] that a Statistical Parametric Mapping (SPM) model, originally developed at the *Wellcome Trust Centre for Neuroimaging* (London) for studying patterns of cerebral activation [Friston *et al.*, 2007], could be a suitable model to describe the pattern of cerebral functional neurodegeneration. Nevertheless SPM suffers the inconveniences of local and univariate approaches, which treat each voxel independently, and hence cannot directly define the functional connectivity in the brain. Subsequently, the sum of abnormal t-values obtained by SPM in regions that were typically hypometabolic in AD has been proposed and used as an AD indicator, with a high accuracy rate [Herholz *et al.*, 2002]. Furthermore a recent study has shown that a voxel-based analysis of ^{18}F FDG-PET increases the diagnostic accuracy and confidence for both AD and FrontoTemporal Dementia (FTD), particularly when findings in a clinical evaluation are not definitive, and physicians are not already highly confident with their clinical diagnosis [Foster *et al.*, 2007].

On the other hand, multivariate approaches consider all voxels of the brain as a single observation. Recent advances in statistical classification and feature extraction techniques have led to a intensive use of those methods. In addition, multivariate approaches are able to surmount the small sample size problem [Duin, 2000]. Most of these multivariate approaches use only a small set of voxels or regions to distinguish between pathological and control images. One of the simplest multivariate approaches for developing a CAD system for AD is the well-known Voxel-As Features (VAF) method [Stoeckel *et al.*, 2001]. This

method separates AD patients and controls by means of a SVM classifier that is trained with all voxels (from SPECT images) with intensity value above a given threshold. Despite its simplicity, this method achieves results similar to other more sophisticated methods. In [López *et al.*, 2009b] a PCA-based feature extraction method is shown. The authors use the Fisher Discrimination Ratio to obtain the most important components obtained by applying a PCA algorithm to the functional images. Then, a Bayesian classifier is used to distinguish between controls and AD patients. Other multivariate approaches use some statistical measures as representation of the brain images. In [Salas-Gonzalez *et al.*, 2009], a reduced map of the brain is computed as the skewness of each m-by-m sliding block of the transaxial slices of the original image. After that, the voxels which present a extreme (very high or very low) Welch's t-statistic between both classes (controls and AD patients) are selected. The mean, standard deviation, skewness and kurtosis are calculated for selected voxels and these measures are chosen as features for three different classifiers: SVM, Decision Trees and Multivariate Normal Model.

12.2. Goals of this Thesis

The main goal of this Thesis is to improve the methods for modeling and analyzing functional tomographic images of the brain and, that way, improve the accuracy of CAD systems for neurodegenerative diseases. Specifically, the goals of this work are divided into two branches:

- On the one hand, the improvement of the CAD systems for neurodegenerative diseases what entails to develop better algorithm for analyzing tomographic images. The work will focus on multivariate approaches, which take a whole brain image as an observation. These systems use supervised statistical learning techniques and do not require that the user interprets the test results. The improvement of these algorithms entails to develop methods that relieve the small sample size problem which previous approaches suffer. The new methods will be derived from the study of two different models. The first one is a classic clustering model, the Gaussian Mixture Model (GMM), which can be used to group or parcel the voxels of functional images. The second one, Partial Least Squares (PLS) technique is a model similar to Principal Component Analysis (PCA) that has been used successfully for regression tasks and can also be used for dimensionality reduction. The development of more accurate CAD systems allows to
-

anticipate the detection of neurodegenerative diseases and thus to improve and extend the life of the patient.

- On the other hand, the implementation of a software tool for managing and analyzing functional images using the methods presented in this Thesis and other previous approaches. The tool's interface should be clear and simple in order to allow that people without computer knowledges may use it.
-

CHAPTER

13

ANALYSIS OF FI BASED ON THE GAUSSIAN MIXTURE MODEL

Clustering or *parcellation* methods are often employed [Jain *et al.*, 1999; Xu & Wunsch, 2005] for the purpose of pattern recognition [Anderberg, 1973], voice detection [Górriz *et al.*, 2006a,b] or image segmentation [Jain & Flynn, 1996]. The basic idea is to group data points, which are similar in some sense, into subsets or parcels. In the case of color images, for instance, these can be contiguous areas of similar color.

In this chapter a new method for analyzing functional images is presented. It is based on the Gaussian Mixture Model, one of the most widely used clustering approach.

13.1. Introduction

Multivariate approaches for developing CAD systems consider as one observation all the voxels in a single scan to make inferences about distributed activation effects. The importance of them is that the effects due to activations, confounding effects and error effects are assessed statistically in terms of effects at each voxel and also interactions among voxels [Frackowiak *et al.*, 2003].

In this context, the classification is usually done by defining feature vectors representing the different functional images and training a classifier with a given set of known samples [Stoeckel *et al.*, 2001, 2004; Fung & Stoeckel, 2007]. After the training process, the classifier is used to distinguish between normal and pathological images. In a straightforward approach the voxel intensities I_n of the functional image are directly used to construct the feature vectors $v = (I_1, \dots, I_N)$, see [Stoeckel *et al.*, 2001, 2004]. Again, even after downsampling the image resolution and applying a brain mask this results in $N \sim 10000$ entries in the feature vectors. Therefore the dimensionality of the feature space is extremely large compared to the number of available training samples (50-100 is a realistic number), which leads to the so-called small sample size problem [Duin, 2000].

Clustering or *parcellation* methods are often employed [Jain *et al.*, 1999; Xu & Wunsch, 2005] for the purpose of data segmentation or compression. The basic idea is to group data points, which are similar in some sense, into subsets or parcels. In the case of color images, for instance, these can be contiguous areas of similar color. Recently, parcels or clusters based on a mixture of Gaussian distributions have been used to quantify the spatial color distribution of images [Goldberger *et al.*, 2006]. Moreover, clustering techniques are successfully applied in various fields like pattern recognition [Anderberg, 1973], speech detection [Górriz *et al.*, 2006a,b], or image segmentation [Jain & Flynn, 1996]. In functional imaging studies, model-based clustering or parcellation has been employed in fMRI analysis for grouping relevant coordinates in the Talaraich space [Newman *et al.*, 2008]. For this task, Activation Likelihood Estimation (ALE) is firstly employed for reducing the list of activation maxima which have one or more maxima in their vicinity and then, these coordinates \mathbf{x}_i with their membership \mathbf{z}_i to each cluster, are subjected to clustering based on finite mixture of probability distributions [Fraley & Raftery, 1998]. The main drawback of this method, derived from the use of ALE, is that hypoperfusion/hypometabolism patterns are not included in the model. The application of model-based image analysis should take into account not only coordinates of activation maxima but the intensity value of each voxel.

In this chapter, a new *parcellation* approach is presented. This approach uses Gaussian Mixtures Models (GMMs) for density estimation of the intensity profile, which allows us to reduce the dimension of the feature vector drastically. We approximate the intensity profile of a functional image by a sum of Gaussians satisfying a maximum likelihood criterion. For this purpose we use the well-known Expectation Maximization (EM) algorithm, due to its simplicity and robustness [Newman *et al.*, 2008], although other maximization methods could be employed as well. Each region of interest (ROI) is then represented by a single Gaussian with a certain center, shape and weight. The feature vectors are constructed by the mean intensities within the different Gaussians, so that the dimensionality of the feature space equals the number of them.

13.2. GMMs for density estimation

GMMs are among the most statistically mature methods for classical clustering see e.g. [Fukunaga, 1990], though they are also used intensively for density estimation [McLachlan & Peel, 2000; Aladjem, 2005]. The basic assumption of GMM for density estimation is that the given data \mathbf{x}_i , $i = 1 \dots N$ are samples drawn from a probability distribution $p(\mathbf{x})$, which is modeled by a sum of k Gaussians

$$p(\mathbf{x}) = \sum_{n=1}^k w_n f_n(\mathbf{x}|\theta_n) \quad (13.1)$$

where $f_n(\mathbf{x}|\theta_n)$ is the density of the Gaussian n with parameter vector θ_n and the w_n are weight factors or mixing proportions with $\sum_n w_n = 1$. The normal distributions $f_n(\mathbf{x}|\theta_n)$ in d dimensions are given by

$$f_n(\mathbf{x}|\theta_n \in \{\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n\}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_n|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_n)^T \boldsymbol{\Sigma}_n^{-1}(\mathbf{x}-\boldsymbol{\mu}_n)} \quad (13.2)$$

with expectation values $\boldsymbol{\mu}_n$ and covariance matrices $\boldsymbol{\Sigma}_n$. Geometrical features of the Gaussians can be varied by parametrization of the covariance matrices $\boldsymbol{\Sigma}_n$ using the eigenvalue decomposition [Banfield & Raftery, 1993]. For our purpose, we assume shape, volume and orientation of the Gaussians variable since the relevant activation areas (ROIs) could be located shapeless and with different sizes across the brain.

13.3. Maximum Likelihood Estimation using the EM algorithm

The Expectation Maximization (EM) algorithm is a parameter estimation method which falls into the general framework of maximum likelihood estimation, and is applied in cases where part of the data can be considered to be incomplete, or “hidden”. The selection of this method for parameter estimation is motivated by certain properties that Dempster et al. demonstrated in [Dempster *et al.*, 1977], in particular at each iteration the loglikelihood of the observed data is guaranteed to be nondecreasing. It is essentially an iterative optimization algorithm that, at least under certain conditions, will converge to parameter values at a local maximum of the likelihood function. The maximum likelihood estimation (MLE) consists of adapting the parameters w_n , $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ in order to maximize the likelihood of a mixture model with k components:

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{i=1}^N p(\mathbf{x}_i|\theta) \quad (13.3)$$

where $\theta = \{\theta_n\}$, for $n = 1, \dots, k$ and $\mathbf{x} = \{\mathbf{x}_i\}$, for $i = 1, \dots, N$, which corresponds to the probability to observe the given samples \mathbf{x}_i , if independent and identically distributed random variables are assumed [Newman *et al.*, 2008; McLachlan & Peel, 2000].

If the data is already grouped into a histogram with B bars at positions \mathbf{x}_j , $j = 1 \dots B$, and with heights h_j , the maximum likelihood estimation can be used in a modified way [McLachlan & Jones, 1988]. In addition, the gray-level of each coordinate is taken into account with the parameter h_j as shown in following section. In that case the total number of observations is given by

$$N = \sum_{j=1}^B h_j \quad (13.4)$$

and the likelihood can be generalized to

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{j=1}^B [p(\mathbf{x}_j|\theta)]^{h_j} \quad (13.5)$$

as there are h_j observations of data points at \mathbf{x}_j . The likelihood to be minimized in equation 13.5 is essentially different from previous approaches [Newman *et al.*,

2008; Fraley & Raftery, 1998] since here, heights h_j are not model parameters to be determined not related to cluster membership.

The maximization is usually achieved by maximizing the log-likelihood instead of the likelihood since both get the maximum at the same values and log-likelihood is easier to determine. The log-likelihood can be formulated as:

$$\ln \mathcal{L}(\theta|\mathbf{x}) = \sum_{j=1}^B h_j \ln \left(\sum_{n=1}^k w_n f_n(\mathbf{x}_j|\theta_n) \right) \quad (13.6)$$

where the histogram heights h_j therefore enter as weight factors. Maximizing the likelihood under the constraint $\sum_{n=1}^k w_n = 1$ is equivalent to finding the maximum of the quantity

$$\mathcal{J} = \sum_{j=1}^B h_j \ln p(\mathbf{x}_j) - \lambda \left(\sum_{n=1}^k w_n - 1 \right), \quad (13.7)$$

where the first term is the log-likelihood and λ is a Lagrange multiplier ensuring the correct normalization.

On the other hand, in the case of “hidden data” \mathbf{z} such that $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ with \mathbf{x} observed and \mathbf{z} unobserved, we can equivalently maximize the so-called complete-data likelihood and find $\hat{\theta}$ such that:

$$\hat{\theta} = \mathbf{arg} \max_{\theta} \mathcal{L}(\theta, z|\mathbf{x}), \quad (13.8)$$

and the complete-data log likelihood may be expressed as [Newman *et al.*, 2008]:

$$\ell(\theta, z|\mathbf{x}) = \sum_{i=1}^N \sum_{n=1}^k z_{in} \log(w_n f_n(\mathbf{x}_i, |\theta_n)), \quad (13.9)$$

where the unknown data $z_i = (z_{i1}, \dots, z_{ik})$ is defined as the membership of the three dimensional vector \mathbf{x}_i in one of the k clusters. The relevant assumptions in equation 13.9 are that the density of an observation \mathbf{x}_i given \mathbf{z}_i is given by $\prod_{n=1}^k f_n(x_i|\theta_k)^{z_{ik}}$ and that each \mathbf{z}_i is independent and identically distributed according to a multinomial distribution of one draw on k categories with probabilities w_1, \dots, w_k [Fraley & Raftery, 1998].

The Maximum Likelihood Estimation in the presence of hidden data can be determined using an Expectation-Maximization algorithm which is described in the next section.

13.4. The EM algorithm

In order to maximize the likelihood, the expectation maximization (EM) algorithm is widely used [Dempster *et al.*, 1977; Moon, 1996] since it is a recursive scheme with interesting monotonic convergence properties. The EM algorithm is a general approach to maximum likelihood in the presence of incomplete data. Starting from an initial guess, the EM algorithm proceeds by alternately estimating the unobservable data z and the unknown parameters θ . Specifically, in the E-step, the algorithm calculates the expected value of the complete data log likelihood with respect to z given x and the current estimate of θ . In the M-step, this expected value is maximized in terms of θ , keeping z fixed as computed in the previous E-step.

The value of z_{in} at the maximum of Eq. 13.9 is the estimated probability that \mathbf{x}_i belongs to cluster n , and the maximum likelihood classification of \mathbf{x}_i is the cluster $n = \operatorname{argmax}_n z_{in}$ for $n = 1, \dots, k$.

Along the same lines as shown for instance in Ref. [Fukunaga, 1990] and using the novel log-likelihood as shown in equation (13.6) we can write down the equations to update the unknown parameters w_n , $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$, where the relations are only modified by a weight factor h_j :

$$w_n = \frac{1}{N} \sum_{j=1}^B h_j q_n(\mathbf{x}_j) \quad (13.10)$$

$$\boldsymbol{\mu}_n = \frac{1}{w_n N} \sum_{j=1}^B h_j q_n(\mathbf{x}_j) \mathbf{x}_j, \quad (13.11)$$

$$\boldsymbol{\Sigma}_n = \frac{1}{w_n N} \sum_{j=1}^B h_j q_n(\mathbf{x}_j) (\mathbf{x}_j - \boldsymbol{\mu}_n)(\mathbf{x}_j - \boldsymbol{\mu}_n)^T \quad (13.12)$$

The posterior probability $q_n(\mathbf{x})$ entering the above equations is defined by

$$q_n(\mathbf{x}) = \frac{w_n f_n(\mathbf{x})}{p(\mathbf{x})} \quad (13.13)$$

Starting with an initial guess for w_n , $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ we can recursively apply Eqs. (13.10), (13.11) and (13.12) until convergence is reached, i.e. the changes in the log-likelihood are smaller than a given threshold.

13.4.1. Some remarks on the EM algorithm

In general the EM algorithm converges monotonically [Dempster *et al.*, 1977], however it is known that the procedure might converge towards a local maximum of the likelihood. Therefore the choice of the initial guess for the unknown parameters w_n , $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ of the Gaussians usually plays an important role. In this sense, in [Fraley & Raftery, 1998] a model-based hierarchical agglomeration method was proposed to yield reasonable clusterings in the absence of any information about a possible clustering inherent in the data. For computational reasons, we prefer to adapt the initialization procedure to the specific data the method is intended for, as shown in the following. Another interesting question about these techniques is the model selection, i.e. to determine the number of parcels inherent to data. Several methods have been proposed to solve model selection problems, i.e. the Bayesian Information Criterion (BIC) [Schwarz, 1978], however the most used one is to apply several models with different pre-defined numbers of components and subsequently choose the best model according to some model selection criterion (computational requirements, required dimensionality reduction, dimension of the feature vector comparable to the number of scans, precision in defining ROIs, etc.)[Górriz *et al.*, 2006a; Goldberger *et al.*, 2006].

13.4.2. Initialization

Since we apply the EM algorithm only to one image, as discussed in section 13.5, it is not crucial to find the global maximum of the likelihood as one and the same GMM configuration is used to extract the features for all images. If the algorithm was applied to each image individually, the initialization would be more important, since the Gaussians might evolve towards qualitatively different configurations corresponding to different local maxima of the likelihood, even if the images are similar.

For our purpose it proved suitable to arrange the k Gaussian centers on a regular grid with $n_x \times n_y \times n_z$ grid points. In order not to “pin” the Gaussian centers at local intensity maxima of the image, it is preferable to initialize the centers at regions of low activations and let them evolve freely. Therefore we use the central subcube with $1/3$ of the side length of the whole brain image, where the intensity is fairly low, and distribute the Gaussian centers regularly across this region. The distances between the grid points are $l_i = L_i/(3n_i)$, with $i = x, y, z$ and L_i being the length of the brain image in the direction i . The weights are initialized with $w_n = 1/k$, and the covariances are chosen to

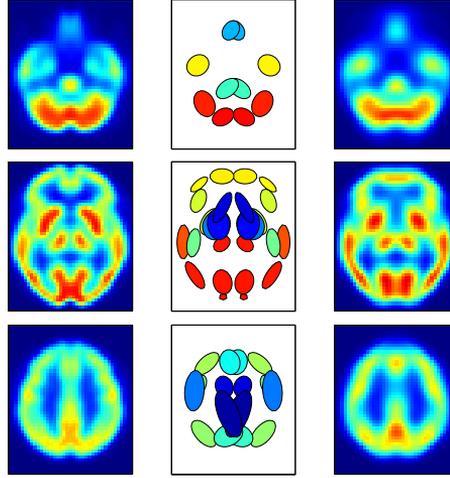


Figure 13.1: Left column: different slices of the original image derived from SPECT image of average normal subjects. Central column: location of the ROIs; the ellipses show the regions of the Gaussians with values larger than 50% of the total height and the colors indicate the intensity of the Gaussians. Right column: reconstructed image from the obtained Gaussians.

be diagonal for the first step of the iteration with elements $3l_x$, $3l_y$ and $3l_z$. This initialization scheme provides a regular and symmetric setup, where the Gaussians initially cover the whole central cube of the brain image and are strongly overlapping.

13.4.3. Model Selection

The model selection criterion is chosen to be data-dependent since given a sufficient number k of Gaussians, all relevant activation areas would be included and/or modeled in them. For the following we use $k = 64$ Gaussians initially arranged on a $4 \times 4 \times 4$ grid. This number allows for a cubic configuration with equal numbers of Gaussian centers in each direction and it turned out to be optimal for the applied initialization scheme: the resulting Gaussians are small and the functional image is reconstructed very well, but still the dimensionality of the feature vectors is of the order of the number of training samples. Other models have been also applied but clearly, they do not meet some of model

selection criteria previously remarked, i.e. using $3^3 = 27$ Gaussians the areas inside the ROIs are larger leading to poorer classification results and a worse reconstruction of the original image; for $10^3 = 1000$ Gaussians the dimensionality of the resulting feature vectors is much larger than the number of training samples.

13.5. Feature Extraction

In this context, feature extraction consists of transforming a functional brain image (i.e. around 500000 voxels once normalized) into a small size vector (preferably less than 100 elements) in order to use it as input for machine learning techniques and to relieve the small sample size problem.

Feature extraction is performed through 3 steps:

- Compute the model image, I_M , which will be modeled with a mixture of Gaussian distributions.
- Run the EM algorithm over I_M in order to determine the underlying mixture of Gaussian distributions.
- Compute feature vectors for all functional images using the mixture of Gaussian distributions obtained in the previous step.

In a first approach, the model image, I_M , could be computed as the average of several control images. We use only controls because the pattern of hypoperfusion/hypometabolism of pathological images is very variable. However, since our main goal is separating controls and pathological images, it could be more interesting to define the model image in another way. A second approach that will result in more effective CAD systems consists of computing I_M as the difference between controls and pathological images:

$$I_M = I_C - I_P \quad (13.14)$$

where I_C and I_P are average of controls and pathological images respectively. Thus, the algorithm will focus only on regions that are important for classification.

Once I_M is computed we run the EM algorithm to obtain the underlying mixture of Gaussian distribution of I_M . A weight value associated with each Gaussian is also calculated. This value indicates the importance of the Gaussian

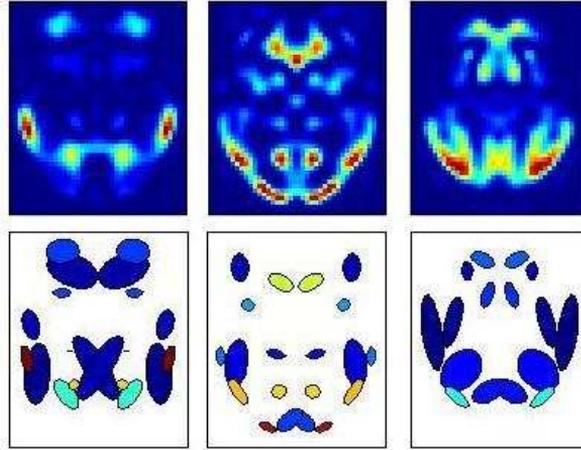


Figure 13.2: ROIs for AD. Top row: Three slices of I_M computed as the difference between controls and AD images. Bottom row: Mixture of Gaussian distributions corresponding to the same image (colors indicate the weight of the Gaussians) [Segovia *et al.*, 2010a]

in the mixture and is an estimation of the importance of the ROI represented by that Gaussian.

Fig. 13.2 shows the image I_M computed as the difference between controls and AD images. The colors indicate the weight of the Gaussians. Note that the regions shown in the figure correspond to those named in the literature as representing regions of Alzheimer's disease [de Leon *et al.*, 1983].

Finally, the mixture of Gaussian distributions obtained is used to calculate a feature vector, $v_{\mathbf{x}}$, for each image.

$$v_{\mathbf{x}} = (c_1, \dots, c_k) \quad (13.15)$$

where c_n stands for the activation of Gaussian n for the image \mathbf{x} and is given by:

$$c_n = h_n \sum_{i=1}^V I(\mathbf{x}_i) f_n(\mathbf{x}_i) \quad (13.16)$$

where $h_n = \frac{w_n}{\sqrt{(2\pi)^3 |\Sigma_n|}}$ and f_n are height and density of Gaussian n respectively.

\mathbf{x}_i are the coordinates of the i th voxel (with $i = 1, \dots, V$) of \mathbf{x} and $I(\mathbf{x}_i)$ is the intensity of the voxel \mathbf{x}_i .

Therefore, the number of features for each image is equal to the number of Gaussians used in the model. However, in order to further reduce the dimensionality of the feature vectors we may ignore the ROIs which are represented by very flat Gaussians. Hence we sort the entries in the feature vectors, Eq. (13.15), according to the height of their Gaussians $w_n/\sqrt{(2\pi)^3|\boldsymbol{\Sigma}_n|}$. Then the feature vectors may be truncated, so that only the most pronounced ROIs are considered.

CHAPTER

14

ANALYSIS OF FI BASED ON THE PARTIAL LEAST SQUARES

Partial Least Squares (PLS) is a wide class of methods for modeling relations between sets of observed variables by means of latent variables. It comprises of regression and classification tasks as well as dimension reduction techniques and modeling tools. The underlying assumption of all PLS methods is that the observed data is generated by a system or process which is driven by a small number of latent (not directly observed or measured) variables [Rosipal & Krämer, 2006].

PLS has received a great amount of attention in scientific community and it is used in many areas including bioinformatics, food research, medicine, pharmacology, social sciences and physiology [Nguyen & Rocke, 2002; Worsley, 1997; Nilsson *et al.*, 1997; Hultand, 1999].

14.1. Model description

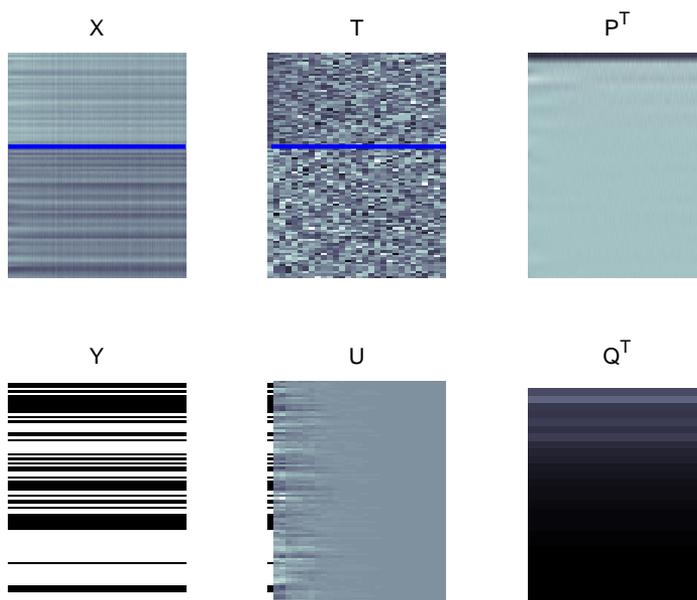
PLS [Wold *et al.*, 1984] is a statistical method for modeling relations among sets of observed variables by means of latent variables. It comprises of regression and classification tasks as well as dimension reduction techniques and modeling tools. The underlying assumption of all PLS methods is that the observed data is generated by a system or process which is driven by a small number of latent (not directly observed or measured) variables. In its general form PLS creates orthogonal score vectors (also called latent vectors or components) by maximizing the covariance among different sets of variables. PLS can be naturally extended to regression problems. The predictor and predicted (response) variables are each considered as a block of variables. PLS then extracts the score vectors which serve as a new predictor representation and regresses the response variables on these new predictors. PLS can be also applied as a discrimination tool and dimension reduction method similar to Principal Component Analysis (PCA). After relevant latent vectors are extracted, an appropriate classifier can be applied.

Mathematically, PLS is a linear algorithm for modeling the relation between two data sets $X \subset \mathbb{R}^N$ and $Y \subset \mathbb{R}^M$. After observing n data samples from each block of variables, PLS decomposes the $n \times N$ matrix of zero-mean variables \mathbf{X} and the $n \times M$ matrix of zero-mean variables Y into the form

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{EY} = \mathbf{UQ}^T + \mathbf{F} \quad (14.1)$$

where the \mathbf{T} , \mathbf{U} are $n \times p$ matrices of the p extracted score vectors (components, latent vectors), the $N \times p$ matrix \mathbf{P} and the $M \times p$ matrix \mathbf{Q} represent matrices of loadings and the $n \times N$ matrix \mathbf{E} and the $n \times M$ matrix \mathbf{F} are the matrices of residuals (or error matrices). The x -scores in \mathbf{T} are linear combinations of the x -variables and can be considered as good summaries of the x -variables. Similarly, the y -scores in \mathbf{U} are linear combinations of the y -variables and can be considered as good summaries of them. Several algorithms have been proposed in the literature to implement the PLS model. In this work, we use the SIMPLS algorithm (see section 14.2). Figure 14.1 illustrates the feature extraction process and the matrices involved in PLS regression of \mathbf{X} and \mathbf{Y} . It is shown how the PLS model accurately matches the given data by means of the extracted score vectors (\mathbf{T} and \mathbf{U}) and loadings (\mathbf{P} and \mathbf{Q}). Note that score matrix corresponding to voxels (\mathbf{T}) contains most information for separating controls and pathological images whereas the loading matrix (\mathbf{P}) is more homogeneous. It is even possible to visually distinguish controls (upper half) and patients (lower half) in the

Figure 14.1: Representation of the matrices involved in the PLS-based method. \mathbf{X} represent a selection of voxels of the images of HVN database and \mathbf{Y} contains the labels (control or AD) assigned to that images.



matrix \mathbf{T} .

The model structures of PLS and PCA are the same in the sense that the data are first transformed into a set of a few intermediate linear latent variables (components) and these new variables are taken into account. Essentially, the difference between PLS and PCA is that the former creates orthogonal weight vectors by maximizing the covariance between elements in \mathbf{X} and \mathbf{Y} . Thus, PLS not only considers the variance of the samples but also considers the class labels. Fisher Discriminant Analysis (FDA) is, in this way, similar to PLS. However, FDA has the limitation that after dimensionality reduction, there are only $c - 1$ meaningful latent variables, where c is the number of classes being considered.

Additionally, when the number of features exceeds the number of samples, the covariance estimates when applying PCA do not have full rank and the weight vectors cannot be extracted.

14.2. SIMPLS algorithm

The SIMPLS algorithm [de Jong, 1993] was proposed by Sijmen de Jong in 1993 as an alternative to the NIPALS algorithm for PLS. The main difference to NIPALS is the kind of deflation. In SIMPLS, no deflation of the centered data matrices \mathbf{X} and \mathbf{Y} is made, but the deflation is carried out for the covariance matrix, or more precisely, the cross-product matrix $\mathbf{S} = \mathbf{X}^T \mathbf{Y}$ between the x -data and y -data [Varmuza & Filzmoser, 2009]. SIMPLS algorithm can be described as follows:

1. initialize $\mathbf{S}_0 = \mathbf{X}^T \mathbf{Y}$ and iterate steps 2 to 8 for $j = 1, \dots, n$
2. if $j = 1$, $\mathbf{S}_j = \mathbf{S}_0$ else, $\mathbf{S}_j = \mathbf{S}_{j-1} - \mathbf{P}_{j-1}(\mathbf{P}_{j-1}^T \mathbf{P}_{j-1})^{-1} \mathbf{P}_{j-1}^T \mathbf{S}_{j-1}$
3. compute w_j as the first singular vector of \mathbf{S}_j
4. $\mathbf{w}_j = \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}$
5. $\mathbf{t}_j = \mathbf{X} \mathbf{w}_j$
6. $\mathbf{t}_j = \frac{\mathbf{t}_j}{\|\mathbf{t}_j\|}$
7. $\mathbf{p}_j = \mathbf{X}_j^T \mathbf{t}_j$
8. $\mathbf{P}_j = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{j-1}]$

The resulting weights \mathbf{w}_j and scores \mathbf{t}_j are stored as columns in the matrix \mathbf{W} and \mathbf{T} respectively. Regression coefficients are calculated as follows:

$$\mathbf{B} = \mathbf{W} \mathbf{T}^T \mathbf{Y} \quad (14.2)$$

14.3. Feature extraction

First, a binary mask is applied to each image in order to remove the voxels that are not part of the brain. Only the voxels that have an intensity above 50 % of maximum intensity in the image computed as the mean of all normal images,

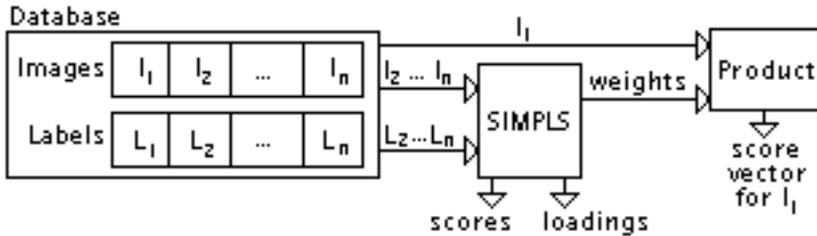


Figure 14.2: Function diagram of the feature extraction method based on PLS. It represents the process followed to calculate the score vector of the first image of the database.

will be considered. Applying this mask to each image leads to a significant reduction of the input space. The initial number of voxels per image ($35 \times 48 \times 40 = 67200$) is reduced to 20638.

Then, score PLS vectors are extracted (matrix \mathbf{T} in Eqn. 14.1) and used as features. In this context, the \mathbf{X} matrix contains the image data: one row per each image and one column per each voxel, i.e. $n - 1$ rows (where n is the size of the database) and 20638 columns. \mathbf{Y} is a vector that contains the labels. In order to avoid biased results, the PLS algorithm is run with all but one image (and their labels) of the database. The score vector for the remaining image is then computed through the weight matrix obtained. This process is repeated for all images of the database. In other words, we applied a leave-one-out methodology to the feature extraction method with the purpose of avoiding that the label of a given image is taken into account to compute its score vector. Fig. 14.2 shows a diagram which describes this process. According to the PLS definition, the weight vector has as many components as images there are in the database minus two and this number will be the size of the feature vectors. However, as same as in PCA, a further reduction of the dimensionality is possible by truncating the feature vectors.

CHAPTER

15

EXPERIMENTS AND RESULTS

Performance of models presented in chapters 13 and 14 has been validated by means of the implementation of several CAD systems for diagnosing Alzheimer's disease. Accuracy, sensitivity and specificity rates obtained are compared with the obtained ones by previous approaches such as PCA [López *et al.*, 2009b] and VAF [Stoeckel *et al.*, 2001]. The statistical measures were estimated using three image databases and two cross-validation techniques.

Table 15.1: Demographic details of the dataset. AD 1 = possible AD, AD 2 = probable AD, AD 3 = certain AD. μ and σ stands for population mean and standard deviation respectively.

	#	Sex (%)		μ	Age	
		M	F		σ	range
NOR	41	73.17	26.83	71.51	7.99	46-85
AD1	30	36.67	63.33	65.86	13.36	23-81
AD2	22	54.55	45.45	67.22	8.25	46-86
AD3	4	0	100.00	76	9.90	69-83

15.1. Image databases

In order to prove that the methods proposed improve the accuracy rates achieved by CAD systems for the diagnosis of neurological diseases, we have used three image databases for Alzheimer’s disease. They are described in the next sections:

15.1.1. HVN database

This database consists of 97 SPECT images collected from a recent study carried out by the “Virgen de las Nieves” hospital in Granada (Spain). The patients were injected with a gamma emitting ^{99m}Tc -ECD radiopharmaceutical and the SPECT raw data was acquired by a three head gamma camera Picker Prism 3000. A total of 180 projections were taken with a 2-degree angular resolution. The images of the brain cross sections were reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter [Ramírez *et al.*, 2008b].

The SPECT images were visually classified by experts of the “Virgen de las Nieves” hospital using 4 different labels: *normal* (NOR) for patients without any symptoms of AD, and *possible AD* (AD1), *probable AD* (AD2) and *certain AD* (AD3) to distinguish among different levels of the presence of typical characteristics for AD. In total, the database consists of 41 NOR, 30 AD1, 22 AD2 and 4 AD3 patients. Table 15.1 shows other demographic details of the database.

15.1.2. CDPC database

The CDPC database consists of 60 images (18 normal and 42 severe AD subjects) provided by the PETCartuja Clinic (Seville, Spain). For PET images, the radiopharmaceutical used was ^{18}F -FDG and the camera was a Siemens CTI ECAT EXACT 47. The post-filter step for PET images consists of a Gaussian filter and an axial filter.

15.1.3. ADNI database

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a cooperative agreement grant that was funded as a public/private partnership with \$40 million from the National Institute on Aging (NIA) and \$20 million from 13 companies in the pharmaceutical industry and two Foundations for a total of \$60 million. Since then, additional companies have joined, bringing the total to 22. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. For more information see <http://adni.loni.ucla.edu/> or <http://www.adni-info.org>.

Data used in this Thesis consist of 403 Fludeoxyglucose (^{18}F -FDG) PET images, acquired from Siemens, General Electric, and Philips PET scanners (Siemens HRRT and BioGraph HiRez scanners were excluded from the primary analysis due to differences in the pattern of FDG uptake). They were collected from the ADNI Laboratory on NeuroImaging (LONI, University of California, Los Angeles). Participants enrollment was conditioned to some eligibility criteria. General inclusion/exclusion criteria was based on measures of disease severity, such as the Mini-Mental State Exam (MMSE) or Clinical Dementia Rating (CDR) were as follows:

- NORMAL control subjects: MMSE scores between 24-30 (inclusive), CDR of 0, non-depressed, non MCI, and non-demented. The age range of normal subjects will be roughly matched to that of MCI and AD subjects.

Table 15.2: Demographic details of the PET images used in this work. μ and σ stand of average and standard deviation respectively

	#	Sex		μ	Age	
		M	F		σ	range
NC	97	60	37	75.97	4.91	62-86
MCI	188	122	66	75.12	7.22	55-89
MCIc	23	18	5	73.97	7.35	57-85
AD	95	57	38	75.72	7.40	55-88

Therefore, there should be minimal enrollment of normals under the age of 70.

- MCI subjects: MMSE scores between 24-30 (inclusive), a memory complaint, have objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of the daily life, and an absence of dementia.
- Mild AD: MMSE scores between 20-26 (inclusive), CDR of 0.5 or 1.0, and meets NINCDS/ADRDA [McKhann *et al.*, 1984] criteria for probable AD.

Therefore, FDG PET data was separated into 3 different classes: Normal Control (NC), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) images. In addition, ADNI sites provide information about MCI who are stable after 2 years follow-up, and those that have converted into AD, i.e. MCI converter (MCIc). The composition of the set of images used in this work is shown in table 15.2.

The PET data were preprocessed according to the procedure described in <http://adni.loni.ucla.edu/about-data-samples/image-data/> in order to remove differences due the scanner used for the acquisition.

15.1.4. Image preprocessing

The images from all databases were spatially normalized using the SPM software [Friston *et al.*, 2007], in order to ensure that the voxels in different images refer to the same anatomical positions in the brain. This step allows us to compare the voxel intensities of the brain images of different subjects. After

the spatial normalization with the SPM software one obtains a $95 \times 69 \times 79$ voxel representation of each subject, where each voxel represents a brain volume of $2,18 \times 2,18 \times 3,56 \text{ mm}^3$

Then intensities were normalized with the maximum intensity I_{max} , which is computed for each image individually by averaging over the 3% of the highest voxel intensities. For ADNI database, I_{max} was computed by averaging the 0.1% of the highest voxel intensities exceeding a threshold. The threshold was fixed to the 10th bin intensity value of a 50-bins intensity histogram, for discarding most low intensity records from outside-brain regions, and preventing image saturation.

15.2. Experiments

In order to evaluate the performance of the feature extraction methods described in chapters 13 and 14, several CAD systems for AD have been implemented using different statistical classifier. Specifically, the classifiers used are:

- Nearest Mean Classifier (NM)
- Fisher Linear Discriminant (FLD)
- Naive Bayes (NB)
- Linear Discriminant Analysis (LDA)
- k Nearest Neighbor Classifier (KNN)
- Decision Trees (DT)
- Random Forests (RF)
- Support Vector Machine (SVM)

With SVM classifier, two kernels function were used: linear and Radial Basis Function (RBF) with parameter $\sigma = 5$.

Five feature extraction methods have been implemented: VAF and PCA used as *baseline*; GMM with two approaches described in section 13.5, that is, using an average of controls (GMM_1) or the difference between controls and pathological images (GMM_2) to define ROIs; and the method based on PLS (Chapter 14). Since all possible combinations of these five methods and nine classifiers are analyzed, in total there are 45 different CAD systems.

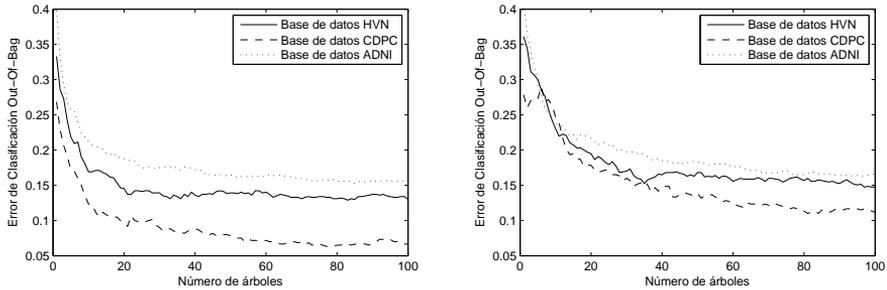


Figure 15.1: Out-Of-Bag error rate obtained by a Random Forests classifier as a function of the number of trees used. *Left*: Out-Of-Bag error rate for GMM-based approach. *Right*: Out-Of-Bag error rate for PLS-based approach.

Experiments with GMM-based systems were run n ($1 \leq n \leq 64$) times using feature vector of size n each time. As it is shown in section 13.5, a model with 64 Gaussians provides feature vectors of 64 components, however, these vectors can be reduced by using only the Gaussians of greater height. Similarly, the PLS-based systems were repeated n times varying the size of the feature vectors. In this case, the maximum size depends on the size of the database ($n = sizeDB - 1$). The system based on VAF and the FLD classifier has not been simulated for computational reason.

Accuracy, sensitivity and specificity rates of the CAD systems were estimated using leave-one-out and k -fold (with $k = 5$ and $k = 10$ [Hastie *et al.*, 2003]) cross-validation techniques. All systems have been trained and validated with three databases of images described in section 15.1.

15.3. Estimation of the classifier parameters

Classifier parameters are automatically adjusted during the training process. However, some parameters, such as the number of trees in a Random Forests classifier must be adjusted manually.

The number of trees in Random Forests, have been adjusted using the Out-Of-Bag error rate, which represents the sum of the classification error of each tree in the forest. Figure 15.1 shows the Out-Of-Bag error rate as a function of the number of trees used. Note that using 50 trees or more, the Out-Of-Bag error rate remains stable. Therefore, using more than 50 trees will not lead to

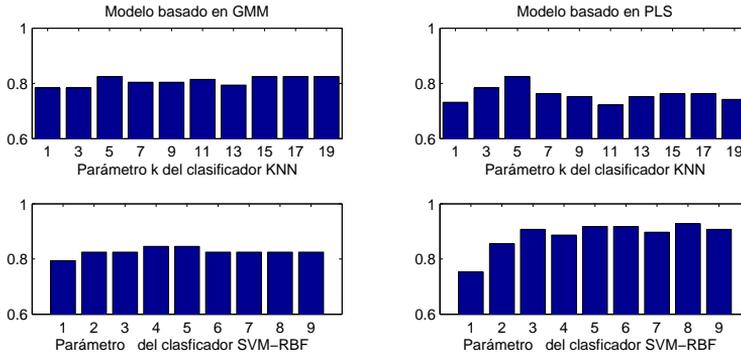


Figure 15.2: Accuracy obtained with a KNN classifiers (top row) and a SVM-RBF (bottom row) as a function of k and σ parameters respectively. These rate are estimated by using 5-fold cross-validation and image from HVN database with GMM approach (left column) and the PLS one (right column)

more accurate systems.

Other parameters such as the value k for KNN classifier or the parameter σ for SVM-RBF classifier were estimated empirically. Figure 15.2 shows a study of the accuracy achieved by both classifiers in function of these parameters. In subsequently experiments we always use $k = 5$ and $\sigma = 5$.

15.4. Results obtained using HVN database

Table 15.3 shows the maximum accuracy rate obtained by each CAD system. Note that, in general, the rates achieved by the two approaches proposed in this Thesis (i.e. GMM- and PLS-based systems) outperform the obtained ones by previous approaches. Highest accuracy rate (96.91 %) is obtained using the PLS-based method and a linear SVM classifier, however, the model called GMM₂ is more stable and achieves accuracy rates around 90 % with most of the classifiers.

15.5. Results obtained using CDPC database

Table 15.3 shows the maximum accuracy rate obtained by each CAD system. The results obtained with this database are significantly better than those

Table 15.3: Accuracy obtained by the developed CAD systems using HVN database. It is shown the maximum rate achieved.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	85.57 %	85.57 %	83.51 %	88.66 %	85.57 %
FLD	-	82.47 %	80.41 %	84.54 %	87.63 %
NB	86.60 %	88.66 %	88.66 %	89.69 %	88.66 %
LDA	-	85.57 %	81.44 %	89.69 %	95.88 %
KNN	80.41 %	82.47 %	84.54 %	85.57 %	84.54 %
DT	83.51 %	79.38 %	82.47 %	90.72 %	86.60 %
RF	84.54 %	88.66 %	89.69 %	89.69 %	87.63 %
SVM (linear)	83.51 %	88.66 %	85.57 %	89.69 %	96.91 %
SVM (RBF)	57.73 %	89.69 %	89.69 %	91.75 %	94.85 %

obtained with the database HVN. In general, classifying high-resolution images, such as PET images, is easier than classifying SPECT images, but in this case, this is also explained by the CDPC database that consists of typical patterns for AD patients and normal controls. On the other hand, the HVN database contains a high proportion of early AD patients as shown in previous sections.

Table 15.4: Accuracy obtained by the developed CAD systems using CDPC database. It is shown the maximum rate achieved.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	86.67 %	90.00 %	75.00 %	88.33 %	95.00 %
FLD	-	98.33 %	98.33 %	98.33 %	98.33 %
NB	88.33 %	95.00 %	73.33 %	93.33 %	95.00 %
LDA	-	98.33 %	96.67 %	96.67 %	88.33 %
KNN	91.67 %	93.33 %	88.33 %	95.00 %	91.67 %
DT	85.00 %	86.67 %	86.67 %	90.00 %	88.33 %
RF	93.33 %	93.33 %	93.33 %	95.00 %	91.67 %
SVM (linear)	96.67 %	96.67 %	96.67 %	96.67 %	91.67 %
SVM (RBF)	70.00 %	96.67 %	96.67 %	98.33 %	91.67 %

It is worth noting the higher accuracy yielded by the FLD classifier. This classifier achieved the highest accuracy rate with all the methods and outperform

the rates achieved by other more sophisticated classifiers such as SVM or RF. This fact proves that the complexity of the classifier plays a key role when training set are small.

15.6. Results obtained using ADNI database

Two experiment have been performed using this database. Since our purpose is to distinguish between healthy subjects and AD patients, first we have trained the CAD systems with only controls and AD images (group 1). Secondly, we have also used MCI images by labeling MCI non converters as controls and MCI converters as AD (group 2). Thus, we can measure the ability of the systems to detect cases of dementia that will progress to AD.

Tables 15.5 and 15.6 show the accuracy rates achieved by the developed CAD systems using the images from group 1 and group 2 respectively.

Table 15.5: Accuracy obtained by the developed CAD systems using images from group 1 of the ADNI database. It is shown the maximum rate achieved.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	81.25 %	80.21 %	80.21 %	85.94 %	83.33 %
FLD	-	81.25 %	72.92 %	79.69 %	82.29 %
NB	80.73 %	76.04 %	82.29 %	88.02 %	85.42 %
LDA	-	84.90 %	80.73 %	86.46 %	86.46 %
KNN	76.04 %	81.25 %	78.12 %	86.98 %	83.85 %
DT	75.00 %	73.44 %	77.08 %	84.38 %	82.29 %
RF	82.81 %	82.81 %	84.38 %	90.62 %	86.98 %
SVM (linear)	80.21 %	87.50 %	82.81 %	87.50 %	85.94 %
SVM (RBF)	50.52 %	85.94 %	84.90 %	89.58 %	85.42 %

Despite of using high-resolution PET images, the rates achieved using images from this database are significantly lower than the obtained with CDPC database. This is due to the composition of ADNI database where images labeled AD correspond to very early AD patients. Thus, distinguish between controls and AD patients is more difficult for ADNI database than for CDPC database. If MCI images are used, classification task become even harder (as prove the low accuracy rates shown in table 15.6) because the high variability of the MCI pattern. Figure 15.3 shows the two first PLS components of all images from AD-

NI database. This figure shows the high variability of the MCI pattern, which makes classification very difficult.

Table 15.6: Accuracy obtained by the developed CAD systems using images from group 2 of the ADNI database. It is shown the maximum rate achieved.

	VAF	PCA	GMM ₁	GMM ₂	PLS
NM	56.08 %	66.00 %	60.30 %	67.00 %	59.80 %
FLD	-	63.28 %	60.55 %	62.03 %	62.78 %
NB	75.68 %	72.21 %	74.94 %	78.41 %	76.18 %
LDA	-	76.43 %	74.94 %	77.17 %	75.68 %
KNN	72.21 %	75.19 %	74.94 %	75.93 %	74.94 %
DT	61.79 %	71.71 %	70.72 %	72.95 %	73.95 %
RF	76.92 %	75.68 %	78.91 %	78.66 %	77.17 %
SVM (linear)	69.98 %	76.67 %	75.68 %	79.40 %	76.67 %
SVM (RBF)	70.72 %	77.17 %	77.67 %	79.40 %	76.67 %

15.7. Discussion

Tables 15.3, 15.4, 15.5 and 15.6 show that the two methods proposed in this Thesis improve, in most cases, the accuracy rates obtained by previous approaches. These results demonstrate that feature selection based on the automatic selection of ROIs or on the Partial Least Squares technique reduces the dimensionality of feature space and preserves relevant information of functional images.

The relevance of the classification results obtained is also confirmed by the ROC curves shown in Fig. 15.4 that measures the trade-off between sensitivity and specificity for varying the number of Gaussians/components.

It is worth noting that CAD systems are reproducing current medical knowledge since they have been trained with samples labeled by clinicians. For this reason, statistical measures reported in this Thesis are an estimation about how a trained system is able to reproduce a medical diagnosis performed by experts. Therefore, some possible errors in the labeling process could render the classification task more difficult, especially for ADNI database where labels were assigned in function of the scores obtained by patients in cognitive tests (as

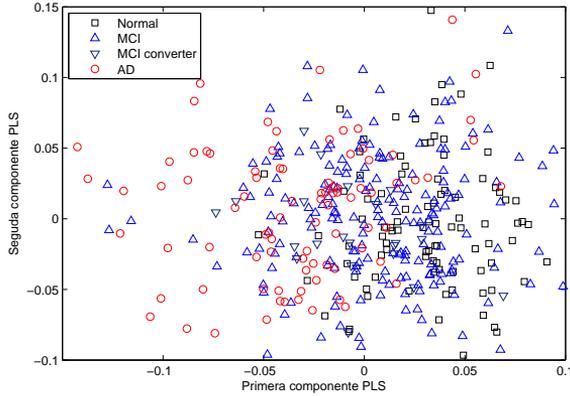


Figure 15.3: Two first PLS components of the 403 images from the ADNI database. Note that distinguishing between MCI and MCI converter is very difficult.

MMSE and CDR) and they do not consider the information provided by the PET images.

In general, the GMM-based method yields better results than the PLS-based one. Parcelation of ROIs using GMM has proven to be an effective approach to extract features. In addition, this method has the advantage that the size of the features vectors does not depend on the number of samples of the database. On the other hand, PLS-based method performs higher accuracy peaks than GMM approach when SPECT images are used (HVN database) and execution time is lower.

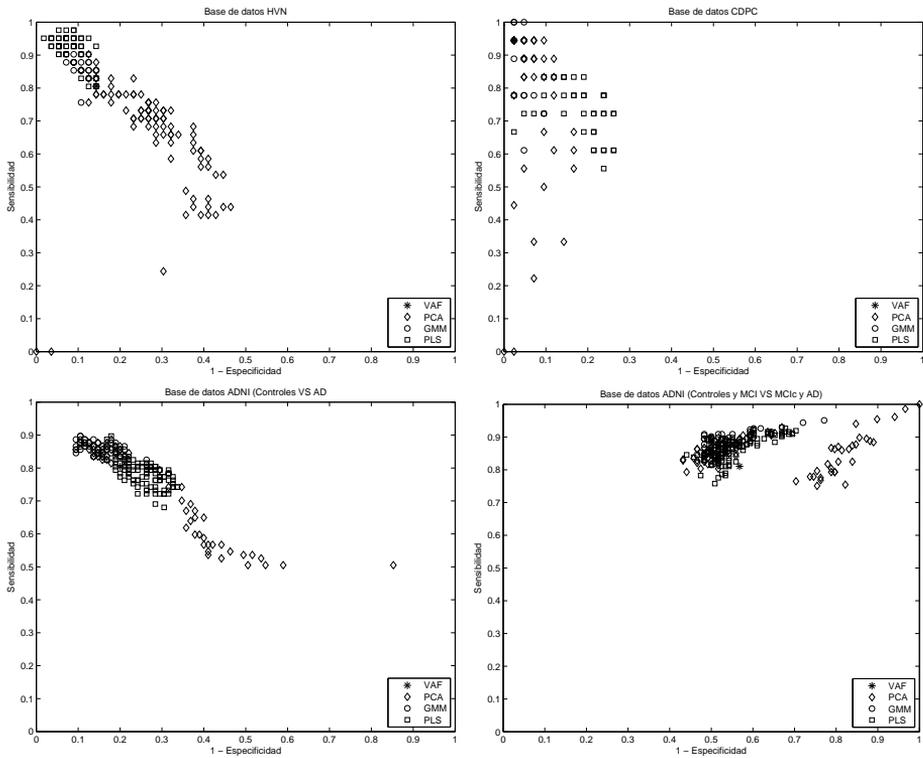


Figure 15.4: ROC curves using the VAF, PCA, GMM_2 and PLS-based methods and the three databases defined on section 15.1. All methods use a linear SVM classifier except GMM_2 that use a SVM-RBF classifier.

CHAPTER

16

TOOL FOR STUDY OF MEDICAL IMAGES

This chapter shows a software tool for computer aided diagnosis of neurological diseases. It allows to manage and analyze functional images using the methods presented in this Thesis and other previous approaches. This tool has an intuitive graphical interface in order to be managed by people without computer knowledge.

16.1. Introduction

The main features of this application are the following:

- Automatic normalization of images, both spatially and in intensity.
- Several forms of displaying images in order to facilitate the comparison. It is also possible to apply different color maps to highlight the differences.
- It is possible to feedback machine learning algorithms with new sets of labeled images.
- Allows simultaneously use of several diagnostic methods, i.e., you can use several methods to obtain several estimated diagnoses of the same image.
- Modular design which allows to easily add new methods.

Since the tool is implemented on Matlab [Attaway, 2009], it may run on Windows, MacOS and Linux. In addition, if it is needed to run the tool on environments not supported by Matlab, it is possible to port the code to C language in a semi-automatic way.

16.2. File formats

Medical images often use specific image formats that allow to store three-dimensional images and encode tag with information about the image such as the patient's name, the image type, the camera and tracer used, etc. In this sense, DICOM and Mayo/Analyze are two specific formats very popular into scientific-medical community. The tool we have developed supports the following file formats:

- **Bipmap (BMP)**. It is a general-purpose format. Thus it can not store additional information about the image and only 2-dimensional images may be stored. In order to store 3-dimensional images is necessary to use multiple bitmap files.
 - **Digital Imaging and Communications in Medicine (DICOM)**. It was designed for handling, storing and transmitting medical images and, nowadays, it has become in a widely used standard. It includes the definition of a file format and a network communication protocol. The communication protocol is an application protocol that uses TCP/IP for
-

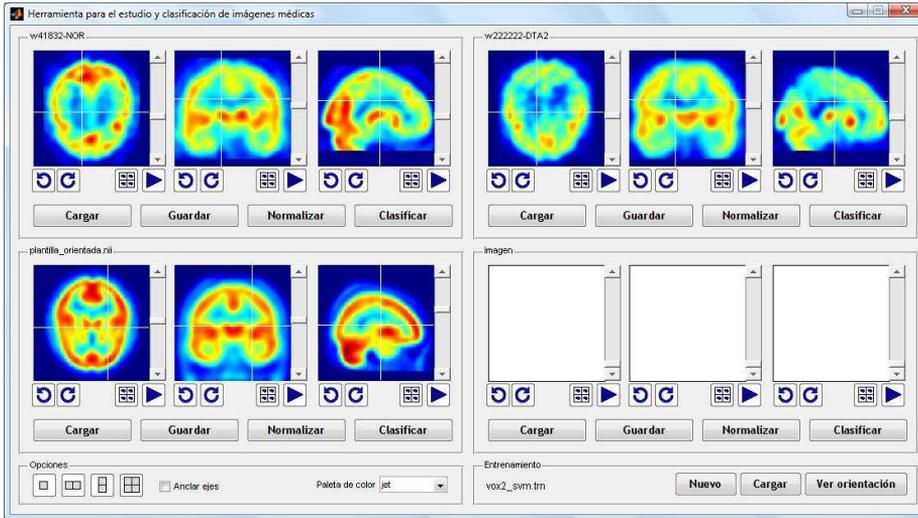


Figure 16.1: Main window of the tool

communication among systems. The first version was developed in 1985 and is currently one of the most popular formats.

- **Mayo/Analyze.** Analyze is a software package developed by Biomedical Imaging Resource (BIR) at Mayo Clinic for visualizing and processing biomedical images. This software introduced a new image format that stores header (information about the image) and image using two different files. This format is used by SPM software.
- **NIfTI-1.** It is an extension of the Mayo/Analyze format proposed by the NIfTI (Neuroimaging Informatics Technology Initiative) initiative. It adds several fields to the header and allows to store header and image in the same file.

16.3. Image Management

Normalization process ensures that two different images are comparable. Obviously, the brain of different patients does not have the same shape and size, thus a direct comparison among them is not possible. In order to ensure that the

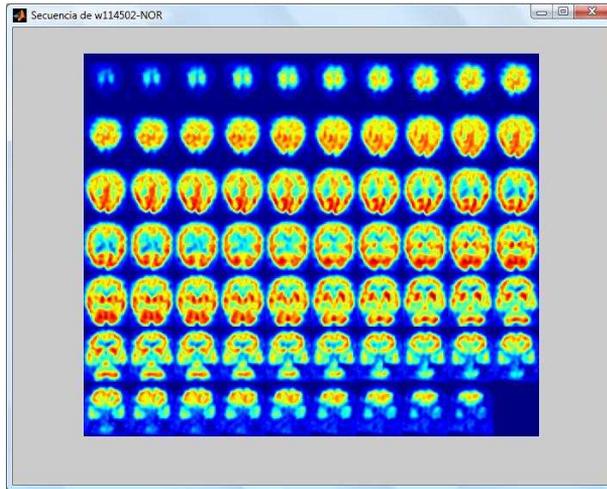


Figure 16.2: Sequence of an SPECT image

same anatomical positions corresponds to the same positions in the image, it is necessary to spatially normalize both images by matching them with a template. On the other hand, small variations in the amount of radiopharmaceutical injected to patients, and the different response to the radiopharmaceutical of each person may cause the intensity levels of different images to vary. Intensity normalization eliminates these differences by equalizing the intensity of voxels with highest intensity and scaling the remaining ones.

Figure 16.1 shows the main window of the tool. Its interface is easy to use and customizable. The main windows allows to visually compare up to four images, each one represented by three slices (one per dimension). Options available for each image are:

- **Normalize.** Perform spatial normalization of the image. During the process it prompts to enter a template.
- **Rotate.** Along to each slice there are two buttons to rotate left or right the image.
- **Sequence.** It shows all possible slices of a given dimension. This representation is often used by clinicians for visually analyzing the images. Figure 16.2 shows an example of this view.

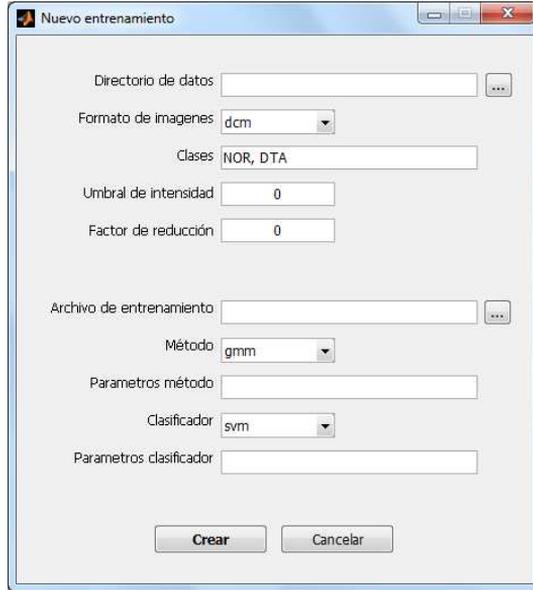


Figure 16.3: Dialog box for creating a new *training*

- **Animation.** It successively displays all slices of the image in a given dimension. The effect is similar to a video stream.
- **Classify.** Perform the automated diagnosis for the image. This feature requires that a training file has been loaded.

16.4. Training and classification

In this context, a *training* is a file that stores a classification function adjusted previously through the training stage of a statistical classifier. Thus, we may classify a given image in real time.

The *training* are created through a specific dialog (figure 16.3) where you must indicate the location of sample images, the classifier type, the feature extraction method to use, etc. The process performs the following steps:

- Read and process the sample images.
-

- Determine the labels of the sample images.
- Extract feature vectors from sample images using the feature extraction method indicated.
- Estimate the parameters of the classification function.
- Save the computed parameters in a file.

Due to the complexity of some feature extraction methods, creating a *training* is a computationally expensive process that can take from a few minutes to several days.

CHAPTER

17

CONCLUSIONS AND FUTURE WORK

This chapter shows the conclusions of the Thesis and highlights the scientific contributions that have been made. These contributions are divided into two groups. On the one hand, the development of new algorithms that improve the precision of CAD systems for neurodegenerative diseases such as Alzheimer's disease and, on the other hand, the implementation of a software tool that brings to hospitals over the past accomplishments in the field of CAD systems. In addition, in this chapter it is enumerated several proposals to continue this work and open new research fields to improve the methods proposed here.

17.1. Conclusions

The main contributions of this Thesis are divided into two branches. On the one hand, the development of new methods for analyzing functional tomographic images, which improve the accuracy of the CAD systems used for the diagnosis of neurodegenerative diseases. And, on the other hand, the implementation of a software tool that brings the latest CAD system improvements, such as the methods described in this work, to hospitals.

Specifically, two methods for analyzing functional images have been developed. In chapter 13 it has been presented a method that makes use of Gaussian mixture models for parceling voxels of functional images and defining, that way, regions of interest. Regions of interest were estimated using an image called *model image*, which can be defined in several ways. Initially we computed *model image* as an average of controls, which leads to define regions of interest as those areas of the brain that are most active in healthy people. Subsequently, in order to improve the results of automatic diagnosis, we decided to define regions of interest as those areas where the mean activity of healthy and sick people differs. Thus, the analysis is focused only on areas that are useful to distinguish between controls and pathological images. Therefore, we computed *model image* as the subtraction between the averages of control and pathological images. Once the *model image* is computed, a Gaussian Mixture algorithm is applied in order to model the regions of interest with a sum of k Gaussians so that the center, variance and weight of the Gaussians stand for position, shape and importance of the regions of interest. The obtained mixture of Gaussian distributions is used to calculate a small size feature vector for each function image. This vector consists of one element for each Gaussian of the mixture, which is calculated as the activation of the entire image for that Gaussian.

The second method proposed is described in chapter 14 and is based on the Partial Least Squares technique. This technique, similar to PCA, transforms two sets of variables into the product of two matrices, called *scores* and *loadings*, according to a criterion of covariance maximization. In our case, the two sets of variables are those formed by the functional images and the labels of these images. Once the transformation is performed, the *score* matrix corresponding to the set of images contains projections of voxels which have been obtained taking into account the labels of the images. These projections called PLS components are used as feature vectors.

The main goals of the proposed methods is to represent a functional image through a small size feature vector which allows to improve the precision of CAD systems for neurodegenerative diseases For this reason, the proposed methods

have been validated by means of four experiments using three functional image databases for the diagnosis of the Alzheimer's disease. Specifically, we have used a database with 97 SPECT images provided by the *Virgen de las Nieves* hospital (Granada, Spain), a database with 60 PET images provided by the *PET Cartuja* center (Sevilla, SPain) and a database with 403 PET images selected from ADNI initiative.

In order to develop the CAD systems, 8 statistical classifiers have been used: Nearest Mean Classifier, Fisher Linear Discriminant, Naive Bayes, Linear Discriminant Analysis, k Nearest Neighbor Classifier, Decision Trees, Random Forests and Support Vector Machine.

The success rates of the resulting CAD systems have been estimated by means of the leave-one-out and k -fold cross-validation techniques. The obtained results have been compared with the obtained ones by previous approaches such as Voxel-As-Features and Principal Component Analysis. They are detailed in chapter 15 and can be summarized as follows:

- PLS-based method achieves an accuracy close to 97% in the diagnosis of AD using SPECT images. That is an improvement of 8 percentage points over the accuracy obtained by the PCA approach and 13 percentage points over the accuracy obtained by VAF approach. In addition, this method is very accurate when we classify PET images but, in this case, the difference with previous approaches is small.
- The GMM-based method is very stable since the results achieved with all the classifiers outperform the obtained ones by previous approaches. Specially, the classification using images from ADNI database are very accurate considering that labeling criterion and database composition make classification very hard. Furthermore, this method achieves a notable performance with a non-linear classifier such as SVM with RBF kernel.

Finally, in chapter 16 it has been shown a software tool that allows to manage and analyze functional images using the methods presented in this Thesis and other previous approaches. Since it has been implemented on Matlab, it is a cross-platform application that may be run on several operating systems. In addition, it has an scalable design that allows us to add new methods in the future.

17.2. Future work

As future lines of the research developed throughout this work, we present the following proposals:

- **Image Fusion.** A drawback of PET images is its higher cost and, for that reason SPECT images are more used despite they provide lower resolution than PET images. On the other hand, Magnetic Resonance Imaging has a high resolution but less ability for showing the metabolic processes. The fusion of SPECT and MRI may become an alternative to PET, economic and capable of producing high-resolution images.
 - **Improvements in normalization process.** Techniques for analyzing functional images have shown a strong dependence on the normalization process. For this reason, developing more accurate normalization techniques will improve CAD systems.
 - **Combination of methods.** The number of Gaussians selected in the GMM-based method should be large enough for obtaining a good representation of the *model image* and small enough for relieving the small sample size problem. A possible alternative consists of generating a model with a large number of Gaussians, getting a high accuracy for reconstruction and then select only a few Gaussians for classification using an alternative method such as PCA or PLS .
 - **Improvement about software tool.** The software tool presented in chapter 16 may be improved in several ways. For example, adding more methods or adapting it to user requirements once the feedback from clinicians have been received.
-

BIBLIOGRAFÍA

- ACTON, PAUL D, & NEWBERG, ANDREW. 2006. Artificial neural network classifier for the diagnosis of Parkinson's disease using [99m Tc]TRODAT-1 and SPECT. *Physics in Medicine and Biology*, **51**(12), 3057.
- ALADJEM, M. 2005. Projection pursuit mixture density estimation. *IEEE Transactions on Signal Processing*, **53**(11), 4376–4383.
- ALMOY, TRYGVE. 1996. A simulation study on comparison of prediction methods when only a few components are relevant. *Computational Statistics and Data Analysis*, **21**(1), 87 – 107.
- ALZHEIMER, ALOIS. 1907. Ueber eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift für Psychiatrie und psychisch-gerichtliche Medizin*, 146–148.
- ANDERBERG, M. R. 1973. *Cluster Analysis for Applications*. Academic Press, New York.
- ANS S. BOSE, L. KOTOULAS, TURKHEIMER, F., PAVESE, N., TAI, P. PICCINI Y.F., HECKEMANN, R.A., & PETROU, M. submitted 2010. Differentiation

- of asymptomatic Huntington's disease from healthy subjects using high order statistics on magnetic resonance images. *Transactions on Medical Imaging*.
- AO MIRANDA, JANAINA MOUR BOKDE, ARUN L.W., BORN, CHRISTINE, HAMPEL, HARALD, & STETTER, MARTIN. 2005. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *NeuroImage*, **28**(4), 980 – 995. Special Section: Social Cognitive Neuroscience.
- AO RICARDO SATO, JO FUJITA, ANDRÉ, THOMAZ, CARLOS EDUARDO, DA GRACA MORAIS MARTIN, MARIA, AO MIRANDA, JANAINA MOUR BRAMMER, MICHAEL JOHN, & JUNIOR, EDSON AMARO. 2009. Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. *NeuroImage*, **46**(1), 105 – 114.
- ARNDT, STEPHAN, CIZADLO, TED, ANDREASEN, NANCY C, HECKEL, DAN, GOLD, SHERRI, & O'LEARY, DANIEL S. 1996. Tests for Comparing Images Based on Randomization and Permutation Methods. *Journal of Cerebral Blood Flow and Metabolism*, **16**, 1271–1279.
- ASHBURNER, J. 2000. Voxel-Based MorphometryThe Methods. *NeuroImage*, **11**(6), 805–821.
- ASHBURNER, J., & FRISTON, K. J. 1999. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, **7**(4), 254–66.
- ASHBURNER, J., NEELIN, P., COLLINS, D.L., EVANS, A.C., & FRISTON, K.J. 1997. Incorporating prior knowledge into image registration. *Neuroimage*, **6**(4), 344–352.
- ATTAWAY, STORMY. 2009. *Matlab: A Practical Introduction to Programming and Problem Solving*. Newton, MA, USA: Butterworth-Heinemann.
- AYACHE, N. 1996. Analyzing 3D Images of the Brain. *NeuroImage*, **4**(3), S34–S35.
- BANFIELD, J., & RAFTERY, A. 1993. Model-based Gaussian and non-Gaussian clustering. *Biometrics*, **49**, 803–821.
- BARKER, M., & RAYENS, W.S. 2003. Partial least squares for discrimination. *Journal of Chemometrics*, **17**(3), 166–173.
-

- BERMEJO-PAREJA, F., BENITO-LEÓN, J., VEGA, S., MEDRANO, M.J., & ROMÁN, G.C. 2008. Incidence and subtypes of dementia in three elderly populations of central Spain. *Journal of the Neurological Sciences*, **264**(1-2), 63 – 72.
- BOSE, SUBRATA K., TURKHEIMER, FEDERICO E., HOWES, OLIVER D., MEHTA, MITUL A., CUNLIFFE, RHIAN, STOKES, PAUL R., & GRASBY, PAUL M. 2008. Classification of schizophrenic patients and healthy controls using [18F] fluorodopa PET imaging. *schizophrenia research*, **106**(2), 1480–155.
- BOURAS, CONSTANTIN, HOF, PATRICK R., GIANNAKOPOULOS, PANTÉLEIMON, MICHEL, JEAN-PIERRE, & MORRISON, JOHN H. 1994. Regional Distribution of Neurofibrillary Tangles and Senile Plaques in the Cerebral Cortex of Elderly Patients: A Quantitative Evaluation of a One-Year Autopsy Population from a Geriatric Hospital. *Cerebral Cortex*, **4**(2), 138–150.
- BOVIS, KEIR, & SINGH, SAMEER. 2002. Classification of Mammographic Breast Density Using a Combined Classifier Paradigm. *Pages 177–180 of: In 4th International Workshop on Digital Mammography.*
- BREIMAN, L., FRIEDMAN, J., OLSHEN, R., & STONE, C. 1984a. *Classification and Regression Trees*. New York: Chapman & Hall.
- BREIMAN, LEO. 1996. Bagging Predictors. *Machine Learning*, **24**(2), 123–140.
- BREIMAN, LEO. 2001. Random Forests. *Mach. Learn.*, **45**(1), 5–32.
- BREIMAN, LEO, FRIEDMAN, JEROME, STONE, CHARLES J., & OLSHEN, R.A. 1984b. *Classification and Regression Trees*. 1 edn. Chapman & Hall/CRC.
- BROOKMEYER, R., JOHNSON, E., ZIEGLER-GRAHAM, K., & ARRIGHI, H.M. 2007. Forecasting the global burden of Alzheimer’s disease. *Alzheimer’s and Dementia*, **3**(3), 186–191.
- BRUYANT, P. P. 2002. Analytic and Iterative Reconstruction Algorithms in SPECT. *The Journal of Nuclear Medicine*, **43**(10), 1343–1358.
- BURGES, C. J. C. 1998. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, **2**(2), 121–167.
- BURNS, A, PHILPOT, M P, COSTA, D C, ELL, P J, & LEVY, R. 1989. The investigation of Alzheimer’s disease with single photon emission tomography. *Journal of Neurology, Neurosurgery and Psychiatry*, **52**(2), 248–253.
-

- CARR, D B, GOATE, A, PHIL, D, & MORRIS, J C. 1997. Current concepts in the pathogenesis of Alzheimer's disease. *The American Journal of Medicine*, **103**(3A), 3S–10S. PMID: 9344401.
- CEKIC, MILOS, CUTLER, SARAH M., STEIN, DONALD G., & WALI, BUSHRA. 2007. Progesterone Improves Acute Recovery after Traumatic Brain Injury in the Aged Rats. *Journal of Neurotrauma*, **24**. 1475–486.
- CHATFIELD, CHRISTOPHER, & COLLINS, ALEXANDER J. 1980. *Introduction to multivariate analysis*. Chapman and Hall LTD.
- CHAVES, R., SALAS-GONZÁLEZ, D., RAMÍREZ, J., GÓRRIZ, J. M., LÓPEZ, M., ÁLVAREZ, I., & SEGOVIA, F. 2010. *Recent Advances in Biomedical Signal Processing*. e-book series. Bentham Science Publishers Ltd. Chap. Preprocessing of Functional Brain Imaging for Computer Aided Diagnosis Systems.
- CHORNOBOY, E. S., CHEN, C. J., MILLER, M. I., MILLER, T. R., & SNYDER, D. L. 1990. An Evaluation of Maximum Likelihood Reconstruction for SPECT. *IEEE Transactions on Medical Imaging*, **9**(1), 99–110.
- CHUI, H., & ZHANG, Q. 1997. *Evaluation of dementia: a systematic study of the usefulness of the American Academy of Neurology's practice parameters*. Tech. rept. American Academy of Neurology.
- C.J.ROSE. 2005. *Statistical Models of Mammographic Texture and Appearance*. Ph.D. thesis, University of Manchester.
- COVER, T. M. 1965. Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. *IEEE Transactions on Electronic Computers*, **14**(3), 326–334.
- CUMMINGS, J. L., VINTERS, H. V., COLE, G. M., & KHACHATURIAN, Z. S. 1998. Alzheimer's disease: etiologies, pathophysiology, cognitive reserve, and treatment opportunities. *Neurology*, **51**(suppl. 1), S2–S17.
- DAVATZIKOS, C., LI, H.H., HERSKOVITS, E., & RESNICK, S.M. 2001. Accuracy and sensitivity of detection of activation foci in the brain via statistical parametric mapping: a study using a PET simulator. *Neuroimage*, **1**, 176–184.
- DE JONG, SIJMEN. 1993. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, **18**(3), 251 – 263.
-

- DE LEON, M.J., FERRIS, S.H., GEORGE, A.E., REISBERG, B., CHRISTMAN, D.R., KRICHEFF, I.I., & WOLF, A.P. 1983. Computed tomography and positron emission transaxial tomography evaluations of normal aging and Alzheimer's disease. *Journal of Cerebral Blood Flow and Metabolism*, **3**(3), 391–394.
- DELBEKE, DOMINIQUE, ROSE, D. MICHAEL, CHAPMAN, WILLIAM C., PINSON, C. WRIGHT, WRIGHT, J. KELLY, BEAUCHAMP, R. DANIEL, SHYR, YU, & LEACH, STEVEN D. 1999. Optimal Interpretation of FDG PET in the Diagnosis, Staging and Management of Pancreatic Carcinoma. *J Nucl Med*, **40**(11), 1784–1791.
- DEMPSTER, A. P., LAIRD, N. M., & RUBIN, D. B. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. Royal Statistical Soc., Ser. B*, **39**(1), 1–38.
- DEVIJVER, PIERRE A., & KITTLER, JOSEF. 1982. *Pattern Recognition: A Statistical Approach*. First edition edn. Prentice Hall.
- DONOGHUE, JOHN P. 2002. Connecting cortex to machines: recent advances in brain interfaces. *Nature Neuroscience*, **5**, 1085–1088.
- DUIN, R. P. W. 2000. Classifiers in almost empty spaces. *Pages 1–7 of: Proceedings 15th International Conference on Pattern Recognition*, vol. 2. IEEE.
- EARY, JANET F. 1999. Nuclear medicine in cancer diagnosis. *The Lancet*, **354**(9181), 853 – 857.
- ENGLISH, R. J., & CHILDS, J. (eds). 1996. *SPECT: Single-Photon Emission Computed Tomography: A Primer*. Society of Nuclear Medicine.
- EVANS, A.C., COLLINS, D.L., MILLS, S.R., BROWN, E.D., KELLY, R.L., & PETERS, T.M. 1993. 3D statistical neuroanatomical models from 305 MRI volumes. *Pages 1813–1817 of: Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, vol. 3.
- FAN, YONG, BATMANGHELICH, NEMATOLLAH, CLARK, CHRIS M., & DAVATZIKOS, CHRISTOS. 2008. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, **39**(4), 1731 – 1743.
- FAWCETT, TOM. 2006. An introduction to ROC analysis. *Pattern Recogn. Lett.*, **27**(8), 861–874.
-

- FERRACCI, FRANCO, CONTE, FERNANDO, GENTILE, MANRICO, CANDEAGO, ROSAMARIA, FOSCOLO, LUCIANO, BENDINI, MATTEO, & FASSETTA, GIULIANO. 1999. Marchiafava-Bignami Disease: Computed Tomographic Scan, ^{99m}Tc HMPAO-SPECT, and FLAIR MRI Findings in a Patient With Subcortical Aphasia, Alexia, Bilateral Agraphia, and Left-handed Deficit of Constructional Ability. *Arch Neurol*, **56**(1), 107–110.
- FILIPPINI, NICOLA, MACINTOSH, BRADLEY J., HOUGH, MORGAN G., GOODWIN, GUY M., FRISONI, GIOVANNI B., SMITH, STEPHEN M., MATTHEWS, PAUL M., BECKMANN, CHRISTIAN F., & MACKAY, CLARE E. 2009. Distinct patterns of brain activity in young carriers of the APOE-E4 allele. *Proceedings of the National Academy of Sciences*, **106**(17), 7209–7214.
- FOSTER, NORMAN L., HEIDEBRINK, JUDITH L., CLARK, CHRISTOPHER M., JAGUST, WILLIAM J., ARNOLD, STEVEN E., BARBAS, NANCY R., DECARLI, CHARLES S., SCOTT TURNER, R., KOEPPE, ROBERT A., HIGDON, ROGER, & MINOSHIMA, SATOSHI. 2007. FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer’s disease. *Brain*, **130**(10), 2616–2635.
- FRACKOWIAK, RICHARD S. J., ASHBURNER, JOHN T., PENNY, WILLIAM D., & ZEKI, SEMIR. 2003. *Human Brain Function, Second Edition*. Academic Press.
- FRALEY, C., & RAFTERY, A.E. 1998. How many clusters? Which Clustering method Answers via model-based cluster analysis. *The computer journal*, **41**(8), 578–588.
- FRANK, ILDIKO E., & FRIEDMAN, JEROME H. 1993. A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, **35**(2), 109–135.
- FREEBOROUGH, P.A., & FOX, N.C. 1998. MR image texture analysis applied to the diagnosis and tracking of Alzheimer’s disease. *Medical Imaging, IEEE Transactions on*, **17**(3), 475–478.
- FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J. P., FRITH, C. D., & FRACKOWIAK, R. S. J. 1994. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, **2**(4), 189–210.
- FRISTON, K. J., ASHBURNER, J., KIEBEL, S. J., NICHOLS, T. E., & PENNY, W. D. 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.
-

- FUKUNAGA, K. 1990. *Introduction to Statistical Pattern Recognition*. Academic Press, New York.
- FUNG, GLENN, & STOECKEL, JONATHAN. 2007. SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. *Knowl. Inf. Syst.*, **11**(2), 243–258.
- GISPERT, J. D., PASCAU, J., REIG, S., GARCÍA-BARRENO, P., & DESCO, M. 2003. Mapas de estadísticos paramétricos (SPM) en medicina nuclear. *Revista Española de Medicina Nuclear*, **22**(1), 43–53.
- GOLDBERGER, J., GORDON, S., & GREENSPAN, H. 2006. Unsupervised image-set clustering using an information theoretic framework. *IEEE Trans. Image Process.*, **15**(2), 449–458.
- GOODMAN, S. N. 1999. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med*, **130**(12), 995–1004.
- GÓRRIZ, J. M., RAMÍREZ, J., SEGURA, J. C., & PUNTONET, C. G. 2006a. An effective cluster-based model for robust speech detection and speech recognition in noisy environments. *J. Acoust. Soc. Am.*, **120**(1), 470–481.
- GÓRRIZ, J. M., RAMÍREZ, J., LANG, E. W., & PUNTONET, C. G. 2006b. Hard C-means clustering for Voice Activity Detection. *Speech Comm.*, **48**(12), 1638–1649.
- GÓRRIZ, J. M., RAMÍREZ, J., LASSL, A., SALAS-GONZALEZ, D., LANG, E. W., PUNTONET, C. G., ÁLVAREZ, I., LÓPEZ, M., & GÓMEZ-RÍO, M. 2008. Automatic Computer Aided Diagnosis Tool using Component-based SVM. *Pages 4392–4395 of: 2008 IEEE Nuclear Science Symposium Conference Record*.
- GOUTIS, CONSTANTINOS. 1996. Partial least squares algorithm yields shrinkage estimators. *The Annals of Statistics*, **24**(2), 816–824.
- HASTIE, T., TIBSHIRANI, R., & FRIEDMAN, J. H. 2003. *The Elements of Statistical Learning*. Corrected edn. Springer.
- HELLAND, INGE S. 2001. Some theoretical aspects of partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, **58**(2), 97 – 107.
-

- HENRY, E.R., & HOFRICHTER, J. 1992. Singular value decomposition: Application to analysis of experimental data. *Pages 129 – 192 of: LUDWIG BRAND, MICHAEL L. JOHNSON (ed), Numerical Computer Methods*, vol. 210. Academic Press.
- HERHOLZ, K., SALMON, E., PERANI, D., BARON, J-C., HOLTHOFF, V., FRÖLICH, L., SCHÖNKNECHT, P., ITO, K., MIELKE, R., KALBE, E., ZÜNDORF, G., DELBEUCK, X., PELATI, O., ANCHISI, D., FAZIO, F., KERRROUCHE, N., DESGRANGES, B., EUSTACHE, F., BEUTHIEN-BAUMANN, B., MENZEL, C., SCHRÖDER, J., KATO, T., ARAHATA, Y., HENZE, M., & HEISS, W-D. 2002. Discrimination between Alzheimer Dementia and Controls by Automated Analysis of Multicenter FDG PET. *NeuroImage*, **17**(1), 302 – 316.
- HERHOLZ, KARL. 1995. FDG PET and Differential Diagnosis of Dementia. *Alzheimer Disease & Associated Disorders*, **9**(1), 6–16.
- HIGDON, R., FOSTER, N. L., KOEPPE, R. A., DECARLI, C. S., JAGUST, W. J., CLARK, C. M., BARBAS, N. R., ARNOLD, S. E., TURNER, R. S., HEIDEBRINK, J. L., & MINOSHIMA, S. 2004. A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer’s disease using FDG-PET imaging. *Statistics in Medicine*, **23**, 315–326.
- HINRICHS, CHRIS, SINGH, VIKAS, XU, GUOFAN, & JOHNSON, STERLING. 2009. MKL for Robust Multi-modality AD Classification. *Pages 786–794 of: MICCAI ’09: Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer-Verlag.
- HOERL, A.E., & KENNARD, R.W. 1970. Ridge regression: bias estimation for nonorthogonal problems. *Technometrics*, **12**, 55 – 67.
- HOFFMAN, E. J., & PHELPS, M. E. 1986. *Positron Emission Tomography and Autoradiography: Principles and Applications for the Brain and Heart*. Chap. Positron emission tomography: principles and quantitation, pages 237–286.
- HOFFMAN, J M, WELSH-BOHMER, K A, HANSON, M, CRAIN, B, HULETTE, C, EARL, N, & COLEMAN, R E. 2000. FDG PET imaging in patients with pathologically verified dementia. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine*, **41**(11), 1920–8. PMID: 11079505.
-

- HOSAKA, KAYO, ISHII, KAZUNARI, SAKAMOTO, SETSU, MORI, TETSUYA, SAKAKI, MASAHIRO, HIRONO, NOBUTSUGU, & MORI, ETSURO. 2002. Voxel-based comparison of regional cerebral glucose metabolism between PSP and corticobasal degeneration. *Journal of the Neurological Sciences*, **199**(1-2), 67 – 71.
- HÖSKULDSSON, A. 1988. PLS regression methods. *Journal of Chemometrics*, **2**(3), 211–228.
- HUDSON, H. M., & LARKIN, R. S. 1994. Accelerated Image Reconstruction Using Ordered Subsets of Projection Data. *IEEE Transactions on Medical Imaging*, **13**(4), 601–609.
- HULLAND, J. 1999. Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal*, **20**(2), 195–204.
- IACOBONI, M., BARON, J.C., FRACKOWIAK, R.S.J., MAZZIOTTA, J.C., & LENZI, G.L. 1999. Emission tomography contribution to clinical neurology. *Clinical Neurophysiology*, **110**, 2–23.
- I.A.ILLAN. 2009 (July). *Análisis en Componentes de Imágenes Funcionales para la Ayuda al Diagnóstico de la Enfermedad de Alzheimer*. Ph.D. thesis, Universidad de Granada.
- IBANEZ, V., PIETRINI, P., ALEXANDER, G. E., FUREY, M. L., TEICHBERG, D., RAJAPAKSE, J. C., RAPOPORT, S. I., SCHAPIRO, M. B., & HORWITZ, B. 1998. Regional glucose metabolic abnormalities are not the result of atrophy in Alzheimer’s disease. *Neurology*, **50**(6), 1585–1593.
- JAIN, A. K., & FLYNN, P. J. 1996. Image segmentation using clustering. *Pages 65–83 of: AHUJA, N., & BOWYER, K. (eds), Advances in Image Understanding: A Festschrift for Azriel Rosenfeld*. IEEE Press, Piscataway, NJ.
- JAIN, A. K., MURTY, M. N., & FLYNN, P. J. 1999. Data clustering: a review. *ACM Comput. Surv.*, **31**(3), 264–323.
- JOACHIMS, T. 1998. Text categorization with Support Vector Machines: Learning with many relevant features. *Pages 137–142 of: Lecture Notes in Computer Science*, vol. 1398.
-

- JOHNSON, STERLING C., SCHMITZ, TAYLOR W., TRIVEDI, MEHUL A., RIES, MICHELE L., TORGERSON, BRITTA M., CARLSSON, CYNTHIA M., ASTHANA, SANJAY, HERMANN, BRUCE P., & SAGER, MARK A. 2006. The Influence of Alzheimer Disease Family History and Apolipoprotein E varepsilon4 on Mesial Temporal Lobe Activation. *J. Neurosci.*, **26**(22), 6069–6076.
- KLÖPPEL, STEFAN, STONNINGTON, CYNTHIA M., CHU, CARLTON, DRAGANSKI, BOGDAN, SCAHILL, RACHAEL I., ROHRER, JONATHAN D., FOX, NICK C., JACK, CLIFFORD R., ASHBURNER, JOHN, & FRACKOWIAK, RICHARD S. J. 2008. Automatic classification of MR scans in Alzheimer's disease. *Brain*, **131**(3), 681–689.
- KOHAVI, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Pages 1137–1143 of: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, vol. 2.
- KOVALEV, V.A., PETROU, M., & BONDAR, Y.S. 1999. Texture anisotropy in 3-D images. *Image Processing, IEEE Transactions on*, **8**(3), 346–360.
- KOVALEV, VASSILI, & PETROU, MARIA. 1996. Multidimensional Co-occurrence Matrices for Object Recognition and Matching. *Graphical Models and Image Processing*, **58**(3), 187–197.
- KOVALEV, VASSILI A., PETROU, MARIA, & SUCKLING, JOHN. 2003. Detection of structural differences between the brains of schizophrenic patients and controls. *Psychiatry Research: Neuroimaging*, **124**(3), 177–189.
- KRÜGER, H., MEESMANN, C., ROHRBACH, E., MÜLLER, J., & HG, H.G. MERTENS. 1990. Panencephalopathic type of Creutzfeldt-Jakob disease with primary extensive involvement of white matter. *European Neurology*, **30**(2), 115–119.
- LAHORTE, P., VANDENBERGHE, S., LAERE, K. VAN, AUDENAERT, K., LE MAHIEU, I., & DIERCKX, R. A. 2000. Assessing the Performance of SPM Analyses of Spect Neuroactivation Studies. *NeuroImage*, **12**(6), 757–764.
- LANGBAUM, JESSICA B.S., CHEN, KEWEI, LEE, WENDY, RESCHKE, COLE, BANDY, DAN, FLEISHER, ADAM S., ALEXANDER, GENE E., FOSTER, NORMAN L., WEINER, MICHAEL W., KOEPPE, ROBERT A., JAGUST, WILLIAM J., & REIMAN, ERIC M. 2009. Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the
-

- Alzheimer's Disease Neuroimaging Initiative (ADNI). *NeuroImage*, **45**(4), 1107–1116.
- LANGE, K., & CARSON, R. 1984. EM reconstruction for emission and transmission tomography. *Journal of Computer Assisted Tomography*, **8**, 306–312.
- LAO, ZHIQIANG, SHEN, DINGGANG, XUE, ZHONG, KARACALI, BILGE, RESNICK, SUSAN M., & DAVATZIKOS, CHRISTOS. 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, **21**(1), 46 – 57.
- LIU, CHENGJUN, & WECHSLER, HARRY. 1998. A Unified Bayesian Framework for Face Recognition. *Pages 151–155 of: In International Conference on Image Processing, ICIP98*.
- LÓPEZ, M., RAMÍREZ, J., GÓRRIZ, J. M., ÁLVAREZ, I., SALAS-GONZALEZ, D., SEGOVIA, F., & CHAVES, R. 2009a. SVM-based CAD System for Early Detection of the Alzheimer's Disease using Kernel PCA and LDA. *Neuroscience Letters*, **464**(4), 233–238.
- LÓPEZ, M., RAMÍREZ, J., GÓRRIZ, J. M., SALAS-GONZALEZ, D., ÁLVAREZ, I., SEGOVIA, F., & PUNTONET, C.G. 2009b. Automatic Tool for the Alzheimer's Disease Diagnosis Using PCA and Bayesian Classification Rules. *IET Electronics Letters*, **45**(8), 389–391.
- MAGNIN, BENOIT, MESROB, LILIA, KINKINGNÉHUN, SERGE, PÉLÉGRINI-ISSAC, MÉLANIE, COLLIOT, OLIVIER, SARAZIN, MARIE, DUBOIS, BRUNO, LEHÉRICY, STÉPHANE, & BENALI, HABIB. 2009. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, **51**(2), 73–83.
- MARDIA, K.V., KENT, J.T., & BIBBY, J.M. 1997. *Multivariate Analysis*. Academic Press.
- MASSY, W.F. 1965. Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, **60**, 234–256.
- McKEITH, IG, BARTHOLOMEW, PH, IRVINE, EM, COOK, J, ADAMS, R, & SIMPSON, AE. 1993. Single photon emission computerised tomography in elderly patients with Alzheimer's disease and multi-infarct dementia. Regional uptake of technetium-labelled HMPAO related to clinical measurements. *The British Journal of Psychiatry*, **163**(5), 597–603.
-

- MCKHANN, GUY, DRACHMAN, DAVID, FOLSTEIN, MARSHALL, KATZMAN, ROBERT, PRICE, DONALD, & STADLAN, EMANUEL M. 1984. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, **34**(7), 939–.
- MCLACHLAN, G., & PEEL, D. 2000. *Finite Mixture Models*. John Wiley and Sons, New York.
- MCLACHLAN, G. J., & JONES, P. N. 1988. Fitting Mixture Models to Grouped and Truncated Data via the EM Algorithm. *Biometrics*, **44**(2), 571–578.
- MILLER, B. L., CUMMINGS, J. L., VILLANUEVA-MEYER, J., BOONE, K., MEHRINGER, C. M., LESSER, I. M., & MENA, I. 1991. Frontal lobe degeneration: Clinical, neuropsychological, and SPECT characteristics. *Neurology*, **41**(9), 1374–1382.
- MILTON, W.J., ATLAS, S.W., LAVI, E., & MOLLMAN, J.E. 1991. Magnetic resonance imaging of Creutzfeldt-Jacob disease. *Annals of Neurology*, **29**(4), 438–440.
- MISSIMER, J., KNORR, U., MAGUIRE, R.P., HERZOG, H., SEITZ, R.J., TELLMAN, L., & LEENDERS, K.L. 1999. On two methods of statistical image analysis. *Hum Brain Mapp*, **8**, 245–258.
- MONTALDI, D, BROOKS, D N, MCCOLL, J H, WYPER, D, PATTERSON, J, BARRON, E, & MCCULLOCH, J. 1990. Measurements of regional cerebral blood flow and cognitive performance in Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, **53**(1), 33–38.
- MOON, T. K. 1996. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, **13**(6), 47–60.
- MORRIS, J.C. 1993. Clinical Dementia Rating. *Neurology*, **43**, 2412–2414.
- MUHIMMAH, IZZATI, & ZWIGGELAAR, REYER. 2006 (October). Mammographic density classification using multiresolution histogram information.
- MUNOZ, YALILE. 2003. Demencia, el reto del presente siglo. *Acta Médica Costarricense*, **45**(2), 42–42.
-

- NEARY, D, SNOWDEN, J S, SHIELDS, R A, BURJAN, A W, NORTHEN, B, MACDERMOTT, N, PRESCOTT, M C, & TESTA, H J. 1987. Single photon emission tomography using ^{99m}Tc -HM-PAO in the investigation of dementia. *Journal of Neurology, Neurosurgery and Psychiatry*, **50**(9), 1101–1109.
- NEWMAN, J., VON CRAMON, D. YVES, & LOHMANN, GABRIELE. 2008. Model-based clustering of meta-analytic functional imaging data. *NeuroImage*, **29**, 177–192.
- NGUYEN, DANH V., & ROCKE, DAVID M. 2002. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*, **18**(1), 39–50.
- NILSSON, JONAS, DE JONG, SIJMEN, & SMILDE, AGE K. 1997. Multiway calibration in 3D QSAR. *Journal of Chemometrics*, **11**(6), 511–524.
- OLIVER, A., FREIXENET, J., & ZWIGGELAAR, R. 2005 (sep.). Automatic classification of breast density. vol. 2.
- ORGANIZATION, WORLD HEALTH (ed). 2006. *Neurological Disorders: Public Health Challenges*. World Health Organization.
- PÉREZ, MIRIAM LÓPEZ. 2010 (July). *Nuevos modelos estadísticos para detección de patrones de hipo/perfusión-metabolismo en imágenes de tomografía funcional cerebral*. Ph.D. thesis, Universidad de Granada.
- PETROU, M. 2010. *Recent Advances in Biomedical Image Processing and Analysis*. Springer-Verlag. Chap. Texture in Biomedical images.
- PETROU, M., KOVALEV, V.A., & REICHENBACH, J.R. 2006. Three-Dimensional Nonlinear Invisible Boundary Detection. *Image Processing, IEEE Transactions on*, **15**(10), 3020–3032.
- PETROUDI, S., KADIR, T., & BRADY, M. 2003 (sep.). Automatic classification of mammographic parenchymal patterns: a statistical approach. vol. 1.
- PETROUDI, STYLIANI, & BRADY, MICHAEL. 2004. Texture Based Mammogram Registration Using Geodesic Interpolating Splines. *Pages 1062–1063 of: MICCAI (2)'04*.
- PHATAK, A., REILLY, P.M., & PENLIDIS, A. 2002. The asymptotic variance of the univariate PLS estimator. *Linear Algebra and its Applications*, **354**, 245–253.
-

- RAMÍREZ, J., GÓRRIZ, J. M., ROMERO, A., LASSL, A., SALAS-GONZALEZ, D., LÓPEZ, M., ALVAREZ, I., GÓMEZ-RÍO, M., & RODRÍGUEZ, A. 2008a. Computer Aided Diagnosis of Alzheimer Type Dementia Combining Support Vector Machines and Discriminant Set of Features. *Accepted in Information Sciences*.
- RAMÍREZ, J., GÓRRIZ, J. M., GÓMEZ-RÍO, M., ROMERO, A., CHAVES, R., LASSL, A., RODRÍGUEZ, A., PUNTONET, C. G., THEIS, F., & LANG, E. 2008b. Effective Emission Tomography Image Reconstruction Algorithms for SPECT Data. *Lecture Notes in Computer Science, Proceedings of the 8th international conference on Computational Science, Part I*, **5101**, 741–748.
- RAUDYS, S.J., & JAIN, A.K. 1991. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans. Pattern Anal. Mach. Intell.*, **13**(3), 252–264.
- RITCHIE, KAREN, & LOVESTONE, SIMON. 2002. The dementias. *The Lancet*, **360**(9347), 1759–1766.
- RODRÍGUEZ, JUAN J., & MAUDES, JESÚS. 2008. Boosting Recombined Weak Classifiers. *Pattern Recognition Letters*, **29**(8), 1049–1059. JCR Impact Factor (2008): 1.559, rank 40/94 in “Computer Science, Artificial Intelligence”, quartile Q2.
- ROSIPAL, ROMAN, & KRÄMER, NICOLE. 2006. Overview and recent advances in partial least squares. *Pages 34–51 of: in ?Subspace, Latent Structure and Feature Selection Techniques?, Lecture Notes in Computer Science*. Springer.
- SALAS-GONZALEZ, D., GÓRRIZ, J. M., RAMÍREZ, J., LASSL, A., & PUNTONET, C.G. 2008a. Improved Gauss-Newton optimisation methods in affine registration of SPECT brain images. *IET Electronics Letters*.
- SALAS-GONZALEZ, D., GÓRRIZ, J. M., RAMÍREZ, J., LASSL, A., & PUNTONET, C. G. 2008b. Improved Gauss-Newton Optimization Methods in Affine Registration of SPECT Brain Images. *IET Electronics Letters*, **44**(22), 1291–1292.
- SALAS-GONZALEZ, D., GÓRRIZ, J. M., RAMÍREZ, J., ÁLVAREZ, I., LÓPEZ, M., SEGOVIA, F., & GÓMEZ-RÍO, M. 2009 (November). Skewness as feature for the diagnosis of Alzheimer’s disease using SPECT images. *In: IEEE International Conference on Image Processing*.
-

- SALAS-GONZALEZ, DIEGO. 2009 (Diciembre). *Nuevos algoritmos de registro y clasificación de imágenes cerebrales PET y SPECT para el diagnóstico de la enfermedad de Alzheimer*. Ph.D. thesis, Universidad de Cadiz.
- SAXENA, P., PAVEL, D. G., QUINTANA, J. C., & HORWITZ, B. 1998. An automatic thresholdbased scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimers disease. *Pages 623–630 of: Medical Image Computing and Computer-Assisted Intervention - MICCAI, Lecture Notes in Computer Science*, vol. 1496.
- SAYEED, A, PETROU, M, SPYROU, N, KADYROV, A, & SPINKS, T. 2002. Diagnostic features of Alzheimer’s disease extracted from PET sinograms. *Physics in Medicine and Biology*, **47**(1), 137.
- SCHAPIRE, R. 2001. *The boosting approach to machine learning: An overview*.
- SCHELTENS, PHILIP. 1999. Early diagnosis of dementia: neuroimaging. *Journal of Neurology*, **246**, 16–20.
- SCHOLKOPF, B., & SMOLA, A. J. 2002. *Learning with Kernels ? Support Vector Machines, Regularization, Optimization and Beyond*. The MIT Press.
- SCHWARZ, G. 1978. Estimating the dimension of a model. *Ann Stat.*, **6**, 461–464.
- SEGOVIA, F., GÓRRIZ, J.M., RAMÍREZ, J., SALAS-GONZÁLEZ, D., ÁLVAREZ, I., LÓPEZ, M., & CHAVES, R. 2010a. Classification of functional brain images using a GMM-based multi-variate approach. *Neuroscience Letters*, **474**(1), 58 – 62.
- SEGOVIA, F., GÓRRIZ, J. M., RAMÍREZ, J., SALAS-GONZÁLEZ, D., ILLÁN, I. A., LÓPEZ, M., & CHAVES, R. 2010b. *Recent Advances in Biomedical Signal Processing*. e-book series. Bentham Science Publishers ltd. Chap. A Comparison Between Univariate And Multivariate Supervised Learning For Classification Of SPECT Images.
- SEGOVIA-MARTINEZ, M., PETROU, M., & CRUM, W. 2001. Texture features that correlate with the mini mental state examination (MMSE) score. vol. 3.
- SHAKHAROVICH, GREGORY, DARRELL, TREVOR, & INDYK, PIOTR. 2006. *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. The MIT Press.
-

- SHEIKH, J.I., & YESAVAGE, J.A. 1986. *Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version*. NY: The Haworth Press.
- SIGNORINI, M., PAULESU, E., FRISTON, K., PERANI, D., LUCIGNANI, G., SOLE, A. DEL, MARTIN, M. DE, STRIANO, G., GRASSI, F., FRACKOWIAK, R.S.J., & FAZIO, F. 1999. Assessment of 18F-FDG PET brain scans in individual patients with statistical parametric mapping. A clinical validation. *NeuroImage*, **9**, 63–80.
- SILVERMAN, D. H., SMALL, G. W., & CHANG, C. Y. 2001a. Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome. *Journal of the American Medical Association*, **286**(17), 2120–2127.
- SILVERMAN, D H, SMALL, G W, CHANG, C Y, LU, C S, ABURTO, M A KUNG DE, CHEN, W, CZERNIN, J, RAPOPORT, S I, PIETRINI, P, ALEXANDER, G E, SCHAPIRO, M B, JAGUST, W J, HOFFMAN, J M, WELSH-BOHMER, K A, ALAVI, A, CLARK, C M, SALMON, E, DE LEON, M J, MIELKE, R, CUMMINGS, J L, KOWELL, A P, GAMBHIR, S S, HOH, C K, & PHELPS, M E. 2001b. Positron emission tomography in evaluation of dementia: Regional brain metabolism and long-term outcome. *JAMA: The Journal of the American Medical Association*, **286**(17), 2120–2127. PMID: 11694153.
- SILVERMAN, DANIEL H S, TRUONG, CO T, KIM, SHANNA K, CHANG, CAROL Y, CHEN, WEI, KOWELL, ARTHUR P, CUMMINGS, JEFFREY L, CZERNIN, JOHANNES, SMALL, GARY W, & PHELPS, MICHAEL E. 2003. Prognostic value of regional cerebral metabolism in patients undergoing dementia evaluation: comparison to a quantifying parameter of subsequent cognitive performance and to prognostic assessment without PET. *Molecular Genetics and Metabolism*, **80**(3), 350–355. PMID: 14680983.
- SMILDE, AGE, BRO, RASMUS, & GELADI, PAUL. 2004. *Multi-way Analysis: Applications in the Chemical Sciences*. Wiley.
- STAMATAKIS, E. A., GLABUS, M. F., WYPER, D. J., BARNES, A., & WILSON, J. T. L. 1999. Validation of Statistical Parametric Mapping (SPM) in Assessing Cerebral Lesions: A Simulation Study. *NeuroImage*, **10**(4), 397 – 407.
-

- STARKSTEIN, S E, MIGLIORELLI, R, TESÓN, A, SABE, L, VÁZQUEZ, S, TURJANSKI, M, ROBINSON, R G, & LEIGUARDA, R. 1994. Specificity of changes in cerebral blood flow in patients with frontal lobe dementia. *Journal of Neurology, Neurosurgery and Psychiatry*, **57**(7), 790–796.
- STOECKEL, J., MALANDAIN, G., MIGNECO, O., KOULIBALY, P. M., ROBERT, P., AYACHE, N., & DARCOURT, J. 2001. Classification of SPECT Images of Normal Subjects versus Images of Alzheimer’s Disease Patients. *Lecture Notes in Computer Science*, **2208**, 666–674.
- STOECKEL, J., AYACHE, N., MALANDAIN, G., KOULIBALY, P. M., EBMEIER, K. P., & DARCOURT, J. 2004. Automatic Classification of SPECT Images of Alzheimer’s Disease Patients and Control Subjects. *Lecture Notes in Computer Science*, **3217**, 654–662.
- STOECKEL, JONATHAN. 2003 (Marzo). *Outils de classification pour l’aide au diagnostic: application à la maladie d’Alzheimer et à d’autres pathologies cérébrales*. Ph.D. thesis, L’Ecole Nationale Supérieure des Mines de Paris.
- TAO, YIMO, LO, SHIH-CHUNG B., FREEDMAN, MATTHEW T., MAKARIOU, ERINI, & XUAN, JIANHUA. 2008. Automatic categorization of mammographic masses using BI-RADS as a guidance. vol. 6915. SPIE.
- THOMPSON, P M, & APOSTOLOVA, L G. 2007. Computational anatomical methods as applied to ageing and dementia. *Br J Radiol*, **80**(2), 78–91.
- TIRABOSCHI, P., HANSEN, L. A., THAL, L. J., & COREY-BLOOM, J. 2004. The importance of neuritic plaques and tangles to the development and evolution of AD. *Neurology*, **62**(11), 1984–1989.
- VAPNIK, V. N. 1982. *Estimation of Dependences Based on Empirical Data*. Springer-Verlag, New York.
- VAPNIK, V. N. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin.
- VAPNIK, V. N. 1998. *Statistical Learning Theory*. John Wiley and Sons, Inc., New York.
- VARDI, Y., SHEPP, L. A., & KAUFMAN, L. 1985. A Statistical Model for Positron Emission Tomography. *Journal of the American Statistical Association*, **80**(389), 8–20.
-

- VARMUZA, KURT, & FILZMOSER, PETER. 2009. *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press.
- VEMURI, PRASHANTHI, GUNTER, JEFFREY L., SENJEM, MATTHEW L., WHITWELL, JENNIFER L., KANTARCI, KEJAL, KNOPMAN, DAVID S., BOEVE, BRADLEY F., PETERSEN, RONALD C., & JR., CLIFFORD R. JACK. 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. *NeuroImage*, **39**(3), 1186 – 1197.
- WANG, GENE-JACK, VOLKOW, NORA D., LEVY, ALEJANDRO V., FELDER, CHRISTOPH A., FOWLER, JOANNA S., PAPPAS, NAOMI R., HITZEMANN, ROBERT J., & WONG, CHRISTOPHER T. 1999. Measuring reproducibility of regional brain metabolic responses to lorazepam using statistical parametric maps. *The Journal of nuclear medicine*, **40**(5), 715–720.
- WANGEN, L. E., & KOWALSK, B. R. 1989. A multiblock partial least squares algorithm for investigating complex chemical systems. *Journal of Chemometrics*, **3**(1), 3–20.
- WENK, G. L. 2003. Neuropathologic changes in Alzheimer's disease. *Journal of Clinical Psychiatry*, **64**(9), 7–10.
- WHITWELL, JENNIFER L., PRZYBELSKI, SCOTT A., WEIGAND, STEPHEN D., KNOPMAN, DAVID S., BOEVE, BRADLEY F., PETERSEN, RONALD C., & JACK, CLIFFORD R. 2007. 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain*, **130**(7), 1777–1786.
- WOLD, H. 1982. Soft modeling: the basic design and some extensions. *Systems under indirect observation*, **2**, 589–591.
- WOLD, S., RUHE, H., WOLD, H., & III, W.J. DUNN. 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverse. *Journal of Scientific and Statistical Computations*, **5**, 735–743.
- WOLPERT, DAVID H. 1992. Stacked Generalization. *Neural Networks*, **5**, 241–259.
- WOODS, R. P. 2000. Spatial Transformation Models. *Chap. 29, pages 465–490 of: BANKMAN, ISAAC N. (ed), Handbook of Medical Imaging*. San Diego: Academic Press.
-

-
- WOODS, R.P., CHERRY, S.R., & MAZIOTA, J.C. 1992. Rapid automated algorithm for aligning and reslicing PET images. *Journal of computer assisted tomography*, **16**(4), 620–633.
- WORSLEY, K. J. 1997. An overview and some new developments in the statistical analysis of PET and fMRI data. *Human Brain Mapping*, **5**, 254–258.
- WORSLEY, K.J., MARRETT, S., NEELIN, P., VANDAL, A.C., FRISTON, K.J., EVANS, A.C., & F. 1996. A Unified Statistical Approach for Determining Significant Signals in Images of Cerebral Activation. *Human Brain Mapping*, **4**, 58–73.
- XU, R., & WUNSCH, D. 2005. Survey of clustering algorithms. *IEEE Trans. Neural Netw.*, **16**(3), 645–678.
- ZHANG, LEI, SAMARAS, D., TOMASI, D., VOLKOW, N., & GOLDSTEIN, R. 2005 (20-25). Machine learning for clinical diagnosis from functional magnetic resonance imaging. *Pages 1211 – 1217 vol. 1 of: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1.
-